

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 29, 1999, 13:01:00 ; Search time 40.75 Seconds  
(Without alignments)  
(46,500 Million cell updates/sec)

Title: US-09-142-524-1

Perfect score: 406  
Sequence: 1 MKYTVAFNQGCPNRRVFIKR.....IASRRVDGIIAAYQNPASWK 80

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	406	100.0	80	1	W27369	Multi-epitope pept
2	327.5	80.7	105	1	W27370	Multi-epitope pept
3	320	78.8	134	1	W27371	Multi-epitope pept
4	144	35.5	47	1	W80353	Sugi allergen prot
5	135	33.3	81	1	W80358	Sugi allergen prot
6	125.5	30.9	47	1	W80352	Sugi allergen prot
7	121.5	29.9	47	1	W80351	Sugi allergen prot
8	120	29.6	47	1	W80357	Sugi allergen prot
9	114	28.1	47	1	W80356	Sugi allergen prot
10	99	24.4	514	1	R53690	Japanese cedar pol
11	99	24.4	514	1	R69792	Japonicum allergen
12	99	24.4	460	1	R69791	Japonicum allergen
13	99	24.4	514	1	R74353	Japanese cedar pol
14	99	24.4	514	1	R81586	Cedar pollen aller
15	99	24.4	514	1	R93599	Japan cedar pollen
16	98	24.4	35	1	W80342	Sugi allergen prot
17	94.5	23.3	33	1	W80340	Sugi allergen prot
18	92.5	22.8	33	1	W80339	Sugi allergen prot
19	91	22.4	33	1	W80338	Sugi allergen prot
20	89	21.9	17	1	R81582	Cedar pollen aller
21	88	21.7	17	1	W80347	Sugi allergen prot
22	88	21.7	47	1	W80350	Sugi allergen prot
23	83	20.4	17	1	R81580	Cedar pollen aller
24	83	20.4	17	1	W80345	Sugi allergen prot
25	81	20.0	15	1	R97884	Japan cedar pollen
26	79	19.5	15	1	W57760	Residues 66-80 of
27	79	19.5	15	1	R97908	Japan cedar pollen
28	79	19.5	15	1	W57764	Residues 186-200 o
29	74	18.2	514	1	W04346	Chamaecyparis obtu
30	74	18.2	375	1	W04344	Chamaecyparis obtu
31	74	18.2	375	1	W04345	Chamaecyparis obtu
32	74	18.2	354	1	W42121	Japanese cypress p
33	74	18.2	514	1	W42122	Japanese cypress p
34	73	18.0	30	1	R45588	Cry j I pollen all
35	73	18.0	367	1	R45577	Jun s I. Antigen
36	73	18.0	15	1	R97907	Japan cedar pollen
37	73	18.0	15	1	W57763	Residues 181-195 o
38	71	17.5	374	1	R31937	Cry j I. Nucleic a
39	71	17.5	36	1	R45586	Cry j I pollen all
40	71	17.5	30	1	R45587	Cry j I pollen all
41	71	17.5	374	1	R45541	Cry j I pollen all
42	71	17.5	374	1	R60166	Japanese cedar pol
43	71	17.5	353	1	R75388	Japanese cedar pol

44 71 17.5 374 1 R82490  
45 71 17.5 353 1 R81587

Cry j I Japanese C  
Cedar pollen aller

## ALIGNMENTS

## RESULT 1

W27369  
W27369 standard; peptide; 80 AA.

AC W27369:

DE 24-MAR-1998 (first entry)

DE Multi-epitope peptide used as immunotherapeutic agent #1.

KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;

OS T-cell epitope region; allergen; lymphocyte; immunoglobulin E.

OS Synthetic.

PN W09732600-AA.

PD 12-SEP-1997.

PE 10-MAR-1997: J00740.

PR 10-MAR-1996: JP-080702.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI Dairiki K, Iwama A, Kuno K, Kume A, Some T;

DR WPI: 97-470495/43.

PT Peptide immunotherapeutic agent to treat allergic diseases -

PT contains multi-epitope peptide containing T cell epitope regions

PT from different allergens

PS Claim 6; Page 31; 58pp; Japanese.

CC The present sequence represents a multi-epitope peptide which is used as

CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2

CC or more different allergens (preferably linked via arginine or lysine

CC dimers), where the T cell epitope regions have a positively index

CC greater than 100 as measured in a patient group responding to the

CC allergen; have at least 70% reactivity with lymphocytes from patients

CC (19E) antibodies from patients responsive to the allergen. The agent can

CC be used to prevent and treat a wide variety of allergic diseases, e.g. by

CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.

Sequence 80 AA;

Query Match 100.0%; Score 406; DB 1; Length 80;

Best Local Similarity 100.0%; Pred. No. 4.2e-48;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKYTVAFNQGCPNRRVFIKRVSNVHIGRRIDIFASKNFMHQNTIGTGRISIKITSGK 60

DB 1 MKYTVAFNQGCPNRRVFIKRVSNVHIGRRIDIFASKNFMHQNTIGTGRISIKITSGK 60

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

DB 61 IASRRVDGIIAAYQNPASWK 80

QY 61 IASRRVDGIIAAYQNPASWK 80

CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 SQ Sequence 105 AA;

Query Match 80.7%; Score 327.5; DB 1; Length 105;  
 Best Local Similarity 67.6%; Pred. No. 2.8e-37;  
 Matches 71; Conservative 1; Mismatches 8; Indels 25; Gaps 2;

QY 1 MKVTAFNPGFNRVETIKRVSNVITIGRRIDIFASKNFHLQKNTIGTGRIS----- 54  
 DB 1 MKVTAFNPGFNRVETIKRVSNVITIGRRIDIFASKNFHLQKNTIGTGRKNNRIMLQ 60  
 QY 54 -LKLSGKIA-----SRVDGIIAAYONPASMK 80  
 61 FAKLTGFTIMGRRLKMPYIAGYKTFDGRVDDGIIAAYONPASMK 105

RESULT 3  
 W27371  
 ID W27371 standard; peptide; 134 AA.  
 AC W27371;  
 DT 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #3.  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 OS Synthetic.  
 PN W09732600-AL.  
 PF 12-SEP-1997.  
 PR 10-MAR-1997; J00740.  
 PR 10-MAR-1996; JP-080702.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Dairiki K, Iwama A, Kume A, Sone T;  
 DR WPI; 97-470495/43.

PT Peptide immunotherapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PS from different allergens  
 PS Claim 6; Page 32; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 SQ Sequence 134 AA;

Query Match 78.8%; Score 320; DB 1; Length 134;  
 Best Local Similarity 53.7%; Pred. No. 4e-36;  
 Matches 72; Conservative 2; Mismatches 6; Indels 54; Gaps 2;

QY 1 MKVTAFNPGFNRVETIKRVSNVITIGRRIDIFASKNFHLQKNTIGTGR----- 52  
 DB 1 MKVTAFNPGFNRVETIKRVSNVITIGRRIDIFASKNFHLQKNTIGTGRKNNRIMLQ 60  
 QY 52 -----ISLKL-----TSKIASRRV 66  
 DB 61 FAKLTGFTIMGRRLKMPYIAGYKTFDGRVDDGIIAAYONPASMK 120  
 QY 67 DGIITAYONPASMK 80  
 |||||

DB 121 DGIITAYONPASMK 134

RESULT 4  
 W80353  
 ID W80353 standard; peptide; 47 AA.  
 AC W80353;  
 DT 11-JAN-1999 (first entry)  
 DE Sugl allergen protein Cryj1 derived epitope for T cells.  
 KW T cell epitope; sugl allergen proteins Cryj1; Cryj2; treatment;  
 KW sugl-pollinosis; allergic reaction; pollen.  
 OS Synthetic.  
 PN J10259198-A.  
 PD 29-SEP-1998.  
 PR 22-DEC-1997; 353448.  
 PR 24-DEC-1996; JP-343441.  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PI (SANY) SANKYO CO LTD.  
 DR WPI; 98-577037/49.  
 PT A linked T cell epitope peptide - used for the treatment of  
 PT sugl-pollinosis  
 PS Claim 10; Page 5; 21pp; Japanese.  
 CC W80353-58 represent epitopes for T cells, derived from the sugl allergen  
 CC proteins Cryj1 (W80353-44, W80350-53 and W80356-58) and Cryj2 (W80345-49  
 CC and W80354-55). The peptides are useful for the treatment of  
 CC sugl-pollinosis, an allergic reaction of the body to pollen.  
 SQ Sequence 47 AA;

Query Match 35.5%; Score 144; DB 1; Length 47;  
 Best Local Similarity 44.3%; Pred. No. 8.9e-13;  
 Matches 35; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MKVTAFNPGFNRVETIKRVSNVITIGRRIDIFASKNFHLQKNTIGTGRISLKLTSK 60  
 DB 13 MKVTAFNPGF-----FASKNFHLQKNT----- 36  
 QY 61 IASRRVDGIIAAYONPASM 79  
 DB 36 -----GIIAAYONPASM 47

RESULT 5  
 W80358  
 ID W80358 standard; peptide; 81 AA.  
 AC W80358;  
 DT 11-JAN-1999 (first entry)  
 DE Sugl allergen protein Cryj1 derived epitope for T cells.  
 KW T cell epitope; sugl allergen proteins Cryj1; Cryj2; treatment;  
 KW sugl-pollinosis; allergic reaction; pollen.  
 OS Synthetic.  
 PN J10259198-A.  
 PD 29-SEP-1998.  
 PR 22-DEC-1997; 353448.  
 PR 24-DEC-1996; JP-343441.  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PI (SANY) SANKYO CO LTD.  
 DR WPI; 98-577037/49.  
 PT A linked T cell epitope peptide - used for the treatment of  
 PT sugl-pollinosis  
 PS Claim 10; Page 5; 21pp; Japanese.  
 CC W80353-58 represent epitopes for T cells, derived from the sugl allergen  
 CC proteins Cryj1 (W80353-44, W80350-53 and W80356-58) and Cryj2 (W80345-49  
 CC and W80354-55). The peptides are useful for the treatment of  
 CC sugl-pollinosis, an allergic reaction of the body to pollen.  
 SQ Sequence 81 AA;

Query Match 33.3%; Score 135; DB 1; Length 81;  
 Best Local Similarity 46.7%; Pred. No. 3e-11;  
 Matches 35; Conservative 2; Mismatches 8; Indels 30; Gaps 2;

QY 1 MKVTAFNPGFNRVETIKRVSNVITIGRRIDIFASKNFHLQKNTIGTGRISLKLTSK 60

DB 14 MKVTAFNFGP-----FASKNFHLOKNT-----KLTSQK 43  
 OY 61 IASRRVDGIITAYON 75  
 DB 44 IASCLNKGIVHANN 58

RESULT 6  
 ID W80352 standard; peptide: 47 AA.  
 AC W80352:  
 DT 11-JAN-1999 (first entry)  
 DE Sugi allergen protein Cryj1 derived epitope for T cells.  
 KW T cell epitope; sugi allergen proteins Cryj1; Cryj2; treatment;  
 KM sugi-pollinosis; allergic reaction; pollen.  
 OS Synthetic.  
 PN J10259198-A.  
 PD 29-SEP-1998.  
 PF 22-DEC-1997: JP-343441.  
 PR 24-DEC-1996: JP-343441.  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA (SANY) SANKYO CO LTD.  
 DR WPI: 98-577037/49.  
 PT A linked T cell epitope peptide - used for the treatment of  
 PT sugi-pollinosis  
 PS Claim 10; Page 5; 21pp; Japanese.  
 CC W80339-58 represent epitopes for T cells, derived from the sugi allergen  
 CC proteins Cryj1 (W80339-44, W80350-53 and W80356-58) and Cryj2 (W80345-49  
 CC and W80354-55). The peptides are useful for the treatment of  
 CC sugi-pollinosis, an allergic reaction of the body to pollen.  
 SQ Sequence 47 AA;

Query Match 30.9%; Score 125.5; DB 1; Length 47;  
 Best Local Similarity 50.8%; Pred. No. 2.9e-10;  
 Matches 32; Conservative 0; Mismatches 0; Indels 31; Gaps 2;  
 OY 1 MKVTAFNFGPNNRVRIFIKRVSNVLIHGRRIDIFASKNFHLOKNTIGTGRIISLKLTSQK 60  
 DB 13 MKVTAFNFGP-----FASKNFHLOKNT-----KLTSQK 41  
 OY 61 IAS 63  
 DB 42 IAS 44

RESULT 7  
 ID W80351 standard; peptide: 47 AA.  
 AC W80351:  
 DT 11-JAN-1999 (first entry)  
 DE Sugi allergen protein Cryj1 derived epitope for T cells.  
 KW T cell epitope; sugi allergen proteins Cryj1; Cryj2; treatment;  
 KM sugi-pollinosis; allergic reaction; pollen.  
 OS Synthetic.  
 PN J10259198-A.  
 PD 29-SEP-1998.  
 PF 22-DEC-1997: JP-343441.  
 PR 24-DEC-1996: JP-343441.  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA (SANY) SANKYO CO LTD.  
 DR WPI: 98-577037/49.  
 PT A linked T cell epitope peptide - used for the treatment of  
 PT sugi-pollinosis  
 PS Claim 10; Page 5; 21pp; Japanese.  
 CC W80339-58 represent epitopes for T cells, derived from the sugi allergen  
 CC proteins Cryj1 (W80339-44, W80350-53 and W80356-58) and Cryj2 (W80345-49  
 CC and W80354-55). The peptides are useful for the treatment of  
 CC sugi-pollinosis, an allergic reaction of the body to pollen.  
 SQ Sequence 47 AA;

Query Match 29.9%; Score 121.5; DB 1; Length 47;  
 Best Local Similarity 58.7%; Pred. No. 1e-09;  
 Matches 27; Conservative 4; Mismatches 4; Indels 11; Gaps 2;

OY 34 FASKNFHLOKNTIGTGRIISLKLTSQKIASRRVDGIITAYONPASW 79  
 DB 13 FASKNFHLOKNT-----MKVT---VAFNFGGIITAYONPASW 47

RESULT 8  
 ID W80357 standard; peptide: 47 AA.  
 AC W80357:  
 DT 11-JAN-1999 (first entry)  
 DE Sugi allergen protein Cryj1 derived epitope for T cells.  
 KW T cell epitope; sugi allergen proteins Cryj1; Cryj2; treatment;  
 KM sugi-pollinosis; allergic reaction; pollen.  
 OS Synthetic.  
 PN J10259198-A.  
 PD 29-SEP-1998.  
 PF 22-DEC-1997: JP-343441.  
 PR 24-DEC-1996: JP-343441.  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA (SANY) SANKYO CO LTD.  
 DR WPI: 98-577037/49.  
 PT A linked T cell epitope peptide - used for the treatment of  
 PT sugi-pollinosis  
 PS Claim 10; Page 5; 21pp; Japanese.  
 CC W80339-58 represent epitopes for T cells, derived from the sugi allergen  
 CC proteins Cryj1 (W80339-44, W80350-53 and W80356-58) and Cryj2 (W80345-49  
 CC and W80354-55). The peptides are useful for the treatment of  
 CC sugi-pollinosis, an allergic reaction of the body to pollen.  
 SQ Sequence 47 AA;

Query Match 29.6%; Score 120; DB 1; Length 47;  
 Best Local Similarity 40.5%; Pred. No. 1.6e-09;  
 Matches 32; Conservative 2; Mismatches 1; Indels 44; Gaps 2;  
 OY 1 MKVTAFNFGPNNRVRIFIKRVSNVLIHGRRIDIFASKNFHLOKNTIGTGRIISLKLTSQK 60  
 DB 1 MKVTAFNFGP-----FASKNFHLOKNT-----KLTSQK 17  
 OY 61 IASRRVDGIITAYONPASW 79  
 DB 18 IAS-CLNKGIVHANNPASW 35

RESULT 9  
 ID W80356 standard; peptide: 47 AA.  
 AC W80356:  
 DT 11-JAN-1999 (first entry)  
 DE Sugi allergen protein Cryj1 derived epitope for T cells.  
 KW T cell epitope; sugi allergen proteins Cryj1; Cryj2; treatment;  
 KM sugi-pollinosis; allergic reaction; pollen.  
 OS Synthetic.  
 PN J10259198-A.  
 PD 29-SEP-1998.  
 PF 22-DEC-1997: JP-343441.  
 PR 24-DEC-1996: JP-343441.  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PA (SANY) SANKYO CO LTD.  
 DR WPI: 98-577037/49.  
 PT A linked T cell epitope peptide - used for the treatment of  
 PT sugi-pollinosis  
 PS Claim 10; Page 5; 21pp; Japanese.  
 CC W80339-58 represent epitopes for T cells, derived from the sugi allergen  
 CC proteins Cryj1 (W80339-44, W80350-53 and W80356-58) and Cryj2 (W80345-49  
 CC and W80354-55). The peptides are useful for the treatment of  
 CC sugi-pollinosis, an allergic reaction of the body to pollen.  
 SQ Sequence 47 AA;

Query Match  
 Best Local Similarity 28.1%; Score 114; DB 1; Length 47;  
 Matches 25; Conservative 0; Mismatches 0; Indels 22; Gaps 1;

34 FASNPHLOKNTIGTRISIKLSKIASNRDGIITAYONPASWK 80  
 1 FASNPHLOKNTIGTRISIKLSKIASNRDGIITAYONPASWK 25

## RESULT 10

R53690  
 ID R53690 standard; Protein; 514 AA.  
 AC R53690;  
 DT 01-FEB-1995 (first entry)  
 DE Japanese cedar pollen allergen Cry j II.  
 KW Cedar pollinosis; diagnostic.  
 OS Cryptomeria japonica.  
 PN WO9411512-A.  
 PD 26-MAY-1994.  
 PE 12-NOV-1993; U11000.  
 PR 12-NOV-1992; US-975179.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Brauer A, Kuo M, Pollock J, Yeung S;  
 DR N-PSDB: 066048.  
 PT Allergen Cry j II protein and fragments from Japanese cedar  
 pollen - used to diagnose, treat and prevent Japanese cedar  
 pollinosis  
 PS Claim 2; Fig 4; 89pp; English.  
 CC The sequence is of a Japanese cedar pollen allergen Cry j  
 II. The protein and its fragments can be used for diagnosis and  
 treatment of Japanese cedar pollinosis and to identify similar  
 CC sequences in other plants.  
 CC See also R53692-6.  
 SQ Sequence 514 AA;

Query Match  
 Best Local Similarity 24.4%; Score 99; DB 1; Length 514;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 IDIFASNPHLOKNTIGTG 49  
 236 IDIFASNPHLOKNTIGTG 254

## RESULT 11

R69792  
 ID R69792 standard; Protein; 514 AA.  
 AC R69792;  
 DT 27-SEP-1995 (first entry)  
 DE Japonicum allergen.  
 KW Japonicum allergen; induced histamine release; antiallergic peptide;  
 OS Japonicum sp.  
 PN WO9502412-A.  
 PD 26-JAN-1995.  
 PE 15-JUL-1994; J01164.  
 PR 16-JUL-1993; JP-177008.  
 PR 01-SEP-1993; JP-217725.  
 PR 07-APR-1994; JP-069336.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Kino K, Kohno Y, Komiyama N, Sone T;  
 DR WPI: 95-067159/09.  
 PT Peptide antiallergic agent - inhibits cross-linking of allergen  
 with IgE antibody  
 PS Example 3; Pages 27-28; 46pp; Japanese.  
 CC Q84045 encodes R69792 Japonicum allergen, from which the antiallergic  
 peptides R69845-R69809 were derived. The peptides ability to inhibit  
 the cross-linking of an allergen, to an IgE antibody can be used in  
 the prevention and treatment of allergic diseases.

SQ Sequence 514 AA;

Query Match  
 Best Local Similarity 24.4%; Score 99; DB 1; Length 514;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 IDIFASNPHLOKNTIGTG 49  
 236 IDIFASNPHLOKNTIGTG 254

## RESULT 12

R69791  
 ID R69791 standard; Protein; 460 AA.  
 AC R69791;  
 DT 27-SEP-1995 (first entry)  
 DE Japonicum allergen residues 55-514.  
 KW Japonicum allergen; residues 55-514; induced histamine release;  
 OS antiallergic peptide; IgE cross-linking inhibition.  
 PN WO9502412-A.  
 PD 26-JAN-1995.  
 PE 15-JUL-1994; J01164.  
 PR 16-JUL-1993; JP-177008.  
 PR 01-SEP-1993; JP-217725.  
 PR 07-APR-1994; JP-069336.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Kino K, Kohno Y, Komiyama N, Sone T;  
 DR WPI: 95-067159/09.  
 PT Peptide antiallergic agent - inhibits cross-linking of allergen  
 with IgE antibody  
 PS Disclosure: Pages 26-27; 46pp; Japanese.  
 CC Q84044 encodes R69791 Japonicum allergen residues 55-514, from  
 which the antiallergic peptides R69845-R69809 were derived.  
 CC The peptides ability to inhibit the cross-linking of an allergen,  
 to an IgE antibody can be used in the prevention and treatment of  
 allergic diseases.  
 SQ Sequence 460 AA;

Query Match  
 Best Local Similarity 24.4%; Score 99; DB 1; Length 460;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 IDIFASNPHLOKNTIGTG 49  
 182 IDIFASNPHLOKNTIGTG 200

## RESULT 13

R74333  
 ID R74333 standard; Protein; 514 AA.  
 AC R74333;  
 DT 01-NOV-1995 (first entry)  
 DE Japanese cedar pollen allergen.  
 KW Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;  
 OS desensitizer.  
 PN WO9502412-A.  
 PD 03-MAY-1995.  
 PE 03-NOV-1994; J08117.  
 PR 05-NOV-1993; JP-299151.  
 PR 20-DEC-1993; JP-344596.  
 PR 27-DEC-1993; JP-346814.  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PI Kurimoto M, Namba M, Torigoe K;  
 DR WPI: 95-195588/26.  
 PT New Japanese cedar pollen allergen polypeptide - and DNA coding for it,  
 useful for treatment and diagnosis of cedar pollen allergy  
 PS Claim 5; Page 26-28; 41pp; English.

CC The gene encoding an allergen of Japanese cedar pollen was isolated  
 CC by PCR amplification using primers based on portions of the allergen  
 CC protein. The gene was used for recombinant allergen production in  
 CC E. coli (vector plasmid pK-223-3).  
 SQ Sequence 514 AA:

Query Match 24.4%: Score 99; DB 1; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 IDIFASKNFHLQKNTIGTG 49  
 ||||||||||||||||  
 Db 236 IDIFASKNFHLQKNTIGTG 254

## RESULT 14

R81586  
 ID R81586 standard; Protein; 514 AA.  
 AC R81586;  
 UA 24-MAY-1996 (first entry)  
 DE Cedar pollen allergen; immunoglobulin E; IgE; T-cell epitope;  
 KW Cedar pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
 OS Cryptomeria japonica.  
 PN EP-700929-A2.  
 PD 13-MAR-1996.  
 PR 08-SEP-1995; 306295.  
 PR 10-SEP-1994; JP-242137.  
 PR 14-JUL-1995; JP-200221.  
 PR 14-JUL-1995; JP-200204.  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PI Hino K, Saito S, Taniguchi Y;  
 DR WPI: 96-140976/15.  
 PR New peptide(s) derived from cedar pollen allergens - activate  
 PR allergen-specific T-cells, but not allergen-specific IgE antibodies,  
 PR used for treating cedar pollinosis.  
 PT Claim 3; Page 29-30; 36pp; English.  
 PS Synthetic peptides based on portions of cedar pollen allergens A  
 CC (R81586) and B (R81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 peptides (R81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (R81573-79) essential  
 CC for T-cell recognition, and homologous peptides (R81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 SQ Sequence 514 AA:

Query Match 24.4%: Score 99; DB 1; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 IDIFASKNFHLQKNTIGTG 49  
 ||||||||||||||||  
 Db 236 IDIFASKNFHLQKNTIGTG 254

## RESULT 15

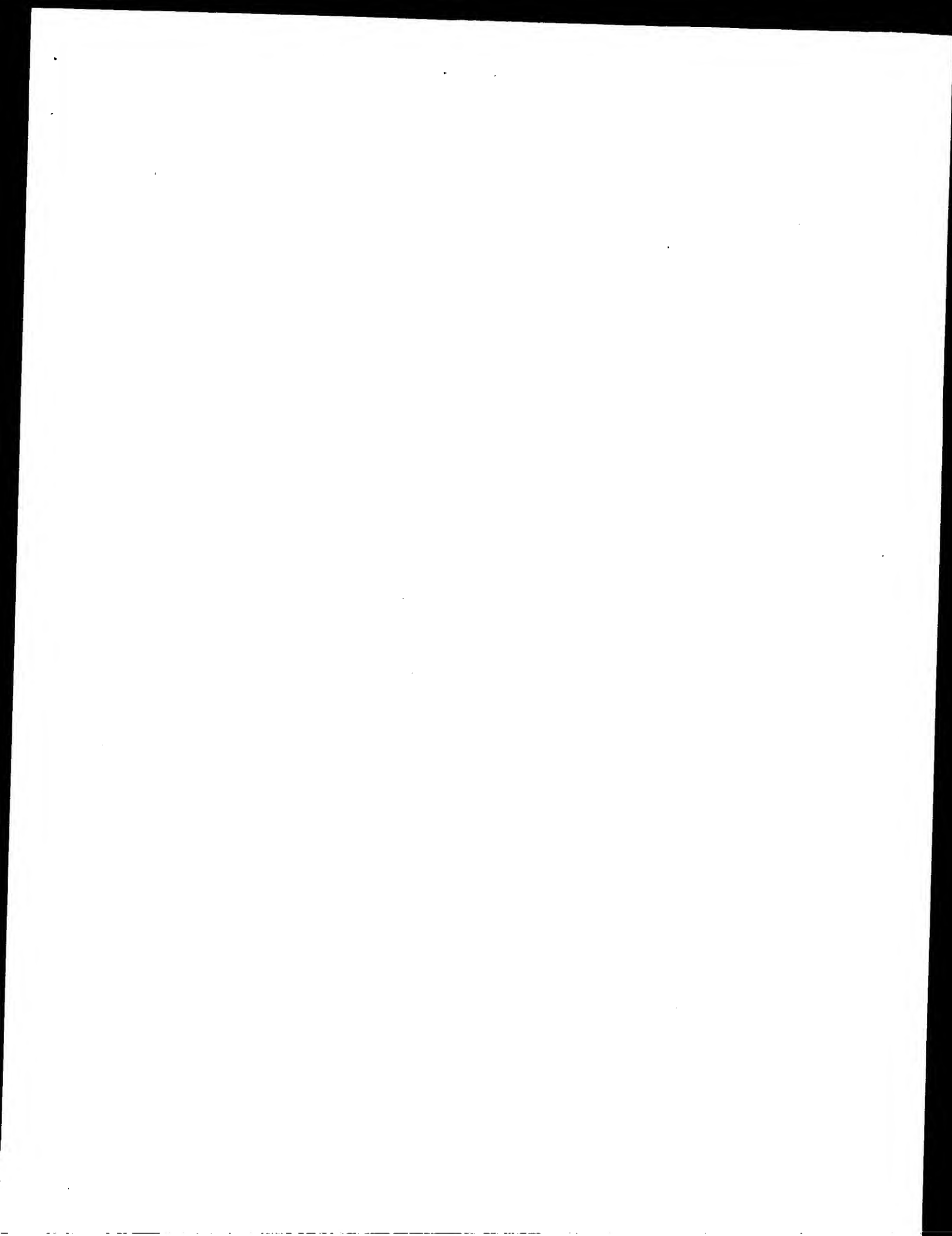
R93599  
 ID R93599 standard; Protein; 514 AA.  
 AC R93599;  
 UA 16-AUG-1996 (first entry)  
 DE Japan cedar pollen Cry j II allergen.  
 KW Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
 OS Sugi pollinosis; diagnosis; treatment.  
 PN Cryptomeria japonica.  
 FH key  
 FT 1..54  
 FT /\*label- sig\_peptide  
 FT 55..514  
 FT /\*label- mat\_protein  
 PN J08047392-A.

PD 20-FEB-1996.  
 PR 07-NOV-1994; 297840.  
 PR 05-NOV-1993; JP-276773.  
 PR 26-MAY-1994; JP-134868.  
 PA (MEIP) MEIJU MILK PROD CO LTD.  
 DR WPI: 96-166249/17.  
 DR N-PSDB; T18102.  
 PR Japan cedar pollen allergen Cry j II epitope - comprises at least  
 PR part of specified 460 amino acid protein  
 PS Claim 1; Page 10-11; 17pp; Japanese.  
 CC R93599 is a Japan cedar pollen Cry j II allergen which is useful  
 CC in the diagnosis, prevention and treatment of Sugi pollinosis,  
 CC the allergic reaction to Japan cedar pollen. Significant regions of  
 CC the allergen were identified using overlapping peptides of the full  
 CC epitope derived from a Cry j II antigen-specific T cell line  
 CC (see R97871-R97960). Amino acids 66-80 (R97884) and 186-200 (R978908)  
 CC of the full mature 460 amino acid allergen are the most allergenic  
 CC of the 90 peptides tested.  
 SQ Sequence 514 AA:

Query Match 24.4%: Score 99; DB 1; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 IDIFASKNFHLQKNTIGTG 49  
 ||||||||||||||||  
 Db 236 IDIFASKNFHLQKNTIGTG 254

Search completed: September 29, 1999, 13:35:10  
 Job time: 2050 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 29, 1999, 13:01:00; Search time 31.02 Seconds  
(without alignments)  
103.328 Million cell updates/sec

Title: US-09-142-524-1

Perfect score: 406  
Sequence: 1 MKVTVAFNQGFRNRYFIKRIK.....IASRRVDGIIAYONPASKR 80

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database: PIR.60:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	99	24.4	514	2	S48730	Cry j II protein -
2	99	24.4	514	2	JC2498	second major aller
3	71	17.5	374	2	JC2124	major allergen Cry
4	71	17.5	374	2	JC2123	major allergen Cry
5	70	17.2	152	1	R3KR18	ribosomal protein
6	70	17.2	152	2	S30393	ribosomal protein
7	70	17.2	152	2	I76666	ribosomal protein
8	66.5	16.4	196	2	S54580	probable membrane
9	66.5	16.3	350	2	S75065	sensory transducti
10	65	16.0	498	2	J02353	glycoprotein E pre
11	64	15.8	423	2	S78174	probable transport
12	62.5	15.4	1280	2	A36466	1-phosphatidylinos
13	62.5	15.3	434	2	S26612	pectate lyase - tr
14	62.5	15.3	173	1	Q0ECP6	fimbrial protein P
15	61.5	15.1	633	2	S62737	topoisomerase I -
16	61	15.0	173	2	D48562	pape fimbrial prot
17	60.5	14.9	110	2	D48562	hypothetical nucle
18	60	14.8	627	2	TVM5V	transforming prote
19	60	14.8	725	2	S60712	hypothetical prote
20	60	14.8	295	2	S60711	band-6-protein - h
21	59.5	14.7	841	2	S34624	band-6-protein - b
22	59.5	14.7	254	2	D65945	glycogen phosphory
23	59.5	14.7	839	2	S53913	ABC transporter, A
24	59	14.5	862	2	S53913	FKB2 protein - yea
25	58.5	14.4	1291	2	S00666	transforming prote
26	58.5	14.4	478	2	S61986	1-phosphatidylinos
27	58.5	14.4	380	2	JC6321	subtilisin-like pr
28	58.5	14.4	380	2	JC6321	CMP-NeuAc:GM3 alph
29	58	14.3	565	1	HMT4DE	hemagglutinin prec
30	58	14.3	152	2	S37496	ribosomal protein
31	58	14.3	578	2	F71844	ABC transporter, A
32	58	14.3	1396	2	G70598	hypothetical prote
33	58	14.3	1396	2	G70598	xanthine dehydroge
34	58	14.3	1070	2	S73093	hypothetical prote
35	57.5	14.2	1290	2	A31317	1-phosphatidylinos
36	57.5	14.2	103	2	E70170	hypothetical prote
37	57	14.0	511	1	XND0U	UTP--glucose-1-pho
38	57	14.0	511	1	I64111	dihydrolipoamide S
39	57	14.0	314	2	JE0127	catechol 2,3-dioxy

40 57 14.0 616 2 JC4084  
41 57 14.0 379 2 S55845  
42 56.5 13.9 167 2 B23093  
43 56.5 13.9 994 2 A47474  
44 56.5 13.9 507137  
45 56.5 13.9 215 2 A48463

prolyl endopeptidase  
3-isopropylmalate  
glycogen phosphory  
NAD+ ADP-ribosyltr  
DNA-directed RNA p  
Ras-like GTP-bind1

## ALIGNMENTS

RESULT 1  
S48730  
Cry j II protein - Japanese cedar  
C:Species: Cryptomeria japonica (Japanese cedar)  
C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Jul-1995  
C:Accession: S48730  
R:Namba, M.; Kurose, M.; Toriige, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.;  
FEBS Lett. 353, 124-128, 1994  
A:Title: Molecular cloning of the second major allergen, Cry j II, from Japanese ceda  
A:Reference number: S48730; MUID:95010777  
A:Accession: S48730  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-514 <NAM>

Query Match 24.4%; Score 99; DB 2; Length 514;  
Best Local Similarity 100.0%; Pred. No. 0.00052;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 IDIFASKNFHLOKNTIGTG 49  
DB 236 IDIFASKNFHLOKNTIGTG 254

RESULT 2  
JC2498  
second major allergen Cry j II precursor - Japanese cedar  
C:Species: Cryptomeria japonica (Japanese cedar)  
C:Date: 16-Mar-1995 #sequence\_revision 26-May-1995 #text\_change 12-Sep-1997  
C:Accession: JC2498; PC2346; A60147  
R:Komiyama, N.; Sone, T.; Shimizu, K.; Morikubo, K.; Kino, K.  
Biochem. Biophys. Res. Commun. 201, 1021-1028, 1994  
A:Title: cDNA cloning and expression of Cry j II, the second major allergen of Japan  
A:Reference number: JC2498; MUID:94271186  
A:Accession: JC2498  
A:Molecule type: mRNA  
A:Residues: 1-514 <KOM>  
A:Cross-references: DDBJ:D29772; NID:g506857; PID:d1006734; PID:g506858  
A:Accession: PC2346  
A:Molecule type: protein  
A:Residues: 52-61 <KO2>  
R:Sakaguchi, M.; Inouye, S.; Tanai, M.; Ando, S.; Usui, M.; Matuhasi, T.  
Allergy 45, 309-312, 1990  
A:Title: Identification of the second major allergen of Japanese cedar pollen.  
A:Reference number: A60147; MUID:90342988  
A:Accession: A60147  
A:Molecule type: protein  
A:Residues: 55-64 <SAK>  
C:Keywords: glycoprotein; pollen  
F:1-54/Domain: signal sequence #status predicted <SIG>  
F:55-460/Product: second major allergen Cry j #status predicted <MAY>  
F:429,460,472/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 24.4%; Score 99; DB 2; Length 514;  
Best Local Similarity 100.0%; Pred. No. 0.00052;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 IDIFASKNFHLOKNTIGTG 49  
DB 236 IDIFASKNFHLOKNTIGTG 254



R.MacMurray, A.J.; Shin, H.S.  
 Mamm. Genome 2, 87-95, 1992  
 A:Title: The murine MHC encodes a mammalian homolog of bacterial ribosomal protein S13.  
 A:Reference number: 157006; MUID:92182530  
 A:Accession: 176666  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-152 <RES>  
 A:Accession: 157006  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-103, 'H', 105-152 <RE2>  
 A:Cross-references: GB:M76762; NID:9198577; PID:9198578  
 C:Genetics:  
 A:Gene: Ke-3  
 A:Introns: 1/3: 34/3; 97/3: 128/2  
 C:Superfamily: Escherichia coli ribosomal protein S13

Query Match 17.2%; Score 70; DB 2; Length 152;  
 Best Local Similarity 30.4%; Pred. No. 0.33;  
 Matches 24; Conservative 7; Mismatches 32; Indels 16; Gaps 2;

QY 12 PNRVRVIRKRVSVNTHGRIDIFASKNFHLQKNTIGGR-----ISLKLTSGLI 61  
 DB 6 PEFQHLRLVLTNIDGRKIAFAITAIK-----GVGRRYAHVYLRKADIDLTRAGEL 59

QY 62 ASRRVDCIAYONPASK 80  
 DB 60 TEDEVETITMONPROCK 78

RESULT 8  
 S54380  
 Probable membrane protein YMR118c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein YW9718.17c  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 08-Jul-1995 #sequence\_revision 19-Oct-1995 #text\_change 21-Nov-1997  
 C:Accession: S54580  
 R:Hunt, S.; Bowman, S.  
 Submitted to the EMBL Data Library, May 1995  
 A:Reference number: S54510  
 A:Accession: S54580  
 A:Molecule type: DNA  
 A:Residues: 1-196 <HUN>  
 A:Cross-references: EMBL:249702; NID:9817859; PID:9817878; MIPS:YMR118c  
 A:Experimental source: strain AB972  
 C:Genetics:  
 A:Map position: 13R  
 C:Keywords: transmembrane protein  
 F:98-114/Domain: transmembrane #status predicted <TM1>  
 F:179-195/Domain: transmembrane #status predicted <TM2>

Query Match 16.4%; Score 66.5; DB 2; Length 196;  
 Best Local Similarity 29.7%; Pred. No. 1.1;  
 Matches 27; Conservative 16; Mismatches 33; Indels 15; Gaps 7;

QY 1 MKVTV--AFNORG-PNRRVFIKRV-----NVTHGRIDIFASKNFH---LQKNTIGR 48  
 DB 1 MKATIGRTSVGVPASVFPVIRISPFILHNTISNG-RMDLF-SKEFHNGRVSKSDLS 58

QY 49 GRRISLKLNSGRIASRRVDGIIAYONPASP 79  
 DB 59 SNKEELIVSOR-KKRPDISPHLTVEPEMSW 88

RESULT 9  
 S75065  
 sensory transduction histidine kinase sll1590 - Synechocystis sp. (strain PCC 6803)  
 N:Alternate names: protein sll1590  
 C:Species: Synechocystis sp.

A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998  
 C:Accession: S75065  
 R:Kaneko, T.; Sato, Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O.; Kato, S.; Okumura, S.; Shimpo, S.; Takeduchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 S:Reference number: S74322; MUID:97061201  
 A:Accession: S75065  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-350 <KAN>  
 A:Cross-references: EMBL:D90910; GB:AB001339; NID:91652956; PID:d1018660; PID:91653001  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Genetics:  
 A:Start codon: GTG

Query Match 16.3%; Score 66; DB 2; Length 350;  
 Best Local Similarity 39.5%; Pred. No. 2.4;  
 Matches 17; Conservative 7; Mismatches 15; Indels 4; Gaps 2;

QY 14 RRVFIKRVSVNTH--GRRIDFASKNFHLQKNTIGGR 52  
 DB 246 QRVFNLLITNAINHSPGRKVEISLTKNNHFOVQIVDEGRGI 288

RESULT 10  
 J02353  
 glycoprotein E precursor - turkey herpesvirus  
 N:Alternate names: ORF 8 protein  
 C:Species: turkey herpesvirus  
 C:Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 09-Sep-1994  
 C:Accession: J02353  
 R:Zelnik, V.; Dattell, R.; Audonnet, J.C.; Smith, G.D.; Riviere, M.; Pastorek, J.;  
 J. Gen. Virol. 74, 2151-2162, 1993  
 A:Title: The complete sequence and gene organization of the short unique region of  
 A:Reference number: J02346  
 A:Accession: J02353  
 A:Molecule type: DNA  
 A:Residues: 1-498 <ZEL>  
 C:Keywords: glycoprotein; transmembrane protein  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-498/Product: glycoprotein E #status predicted <MAT>  
 F:396-412/Domain: transmembrane #status predicted <TM1>  
 F:44,60,133,148,370/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 16.0%; Score 65; DB 2; Length 498;  
 Best Local Similarity 31.6%; Pred. No. 4.7;  
 Matches 18; Conservative 10; Mismatches 25; Indels 4; Gaps 2;

QY 18 IKRVSVNTHGRIDIFASKNFHLQKNT-IGTGRISLKLTSKRIASRRVDGIIAY 73  
 DB 92 ILRVDFVNVSGSLDRIASQFHYIRVITIGTGKRLTFN--ATSOIAGVYTRY 145

RESULT 11  
 S78174  
 Probable transport protein yojw - Reclinomonas americana (ATCC 50394) mitochondriac  
 C:Species: mitochondrion Reclinomonas americana  
 A:Variety: ATCC 50394  
 C:Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 17-Mar-1999  
 C:Accession: S78174  
 R:Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C  
 Nature 387, 493-497, 1997  
 A:Title: An ancestral mitochondrial DNA resembling a eubacterial genome in miniat  
 A:Reference number: S78127; MUID:97311393  
 A:Accession: S78174  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-222 <LAN>

Qy	18	IKRVSNIYH--GRID--IFASKNFHÖKNTI-----CTGRISIKTISGRIASNR	65	12; Gaps
Db	3	AKIKNLHTHTTIGIRSNKIIFONINFSLEKSLIIIOGNSGSGTKTLTKITISGLPPKQ	62	
Qy	66	VDGIT 70		
Db	63	GDIII 67		

## A36466

1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase (EC 3.1.4.11) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 15-Feb-1991 #sequence  
 C:Accession: A36466  
 C:Revision: 15-Feb-1991 #text-change 12-Feb-1999  
 M:Inrgers, W.H.; Dione, C.A.; Kaplow, J.; Mudd, R.; Friesel, R.; Zilberstein, A.; Schid  
 Mol. Cell. Biol. 10, 4770-4777, 1990  
 A:Title: Characterization and cDNA cloning of phospholipase C-gamma, a major substrate f  
 A:Reference number: A36466; MUID:90353993  
 A:Accession: A36466  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 11290 <B0R>  
 A:Cross-references: GB:M34667; NID:g190037; PID:g190038  
 A:Gene: GDB:PLCG1; PLC1  
 A:Cross-references: GDB:120299; OMIM:172420  
 A:Map position: 20q12-20q13.1  
 C:Superfamily: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase II; 1-phosphati  
 C:Keywords: phosphoprotein; pleckstrin repeat homology; SH2 homology; SH3 homology  
 F:332-466/Domain: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X hom  
 F:550-657/Domain: SH2 homology <SH21>  
 F:668-736/Domain: SH2 homology <SH12>  
 F:798-846/Domain: SH3 homology <SH3>  
 F:952-1073/Domain: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X hom

Query Match	15.48;	Score 62.5;	DB 2;	Length 1290;
Best Local Similarity	28.18;	Pred. No. 27;		
Matches	18;	Conservative	9;	Mismatches 34.

	34	Angels	3	Gaps	1
QY	5	VAENOPGNRRVFEIKRVSNVLIHGRRIDFASKNNHLO--	KNLTGTGRISILKITSKI	61	
Db	883	IAIREGKNNRRLFEVSIASVAHMSLDVAADSOEHLQDMVKIR	EVATADARLTGKRI	942	
QY	62	ASRR	65		
Db	943	MERR	946		

RESULT 13  
S29612  
pectate lyase - trumpet lily  
C:Species: Lillium longiflorum (trumpet lily)

C:\Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 18-Sep-1998  
A:Accession: S29612  
R:Kim, S., Finkel, D.J., An, G.  
Submitted to the EMBL Data Library, October 1992  
A:Description: Abundant patterns of lily pollen cDNAs: characterization of three p  
A:Reference number: S29611  
A:Accession: S29612  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-434 <K1M>  
A:Cross-references: EMBL:Z17328, NTD:G19450; PID:G19451  
C:\Superfamily: pectate lyase LPT59

	Indels	Gaps
Qy 1 MKVYAFNFGFNN-----RRVIRKVSNTIIGRIDYASKNFHLQKTTIGTR- 52	33	19
Db 206 KQVYAFNFGFGLGLORPFCRYGVEHVNDYTHWIMYAVGSOH-----PLISQGNRY 351		
Qy 52 ISLKITSSKIASRRVDGIIAAYQNDASW 79		
Db 352 IAPIHEAKVEYTKR-----DYAEPAEW 373		

अध्याय ४

```

fimbrial protein pape precursor - Escherichia coli
1QCEP
N:Alternate names: pape protein
C:Species: Escherichia coli

```

C;Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 29-May-1998  
C;Accession: A2134; A27743; S25222; S25227; S16401; S16406  
R;Lindberg, F.; Lund, B.; Normark, S.  
Proc Natl Acad Sci USA 85:1111-1115 (1988)

Author: Matt. Acad. Sci. U.S.A. 83, 1891-1895, 1986  
 Article: Gene products specifying adhesion of uropathogenic *Escherichia coli* are minor  
 Reference number: A25134; MUID:86199403  
 Accession: A25134

A: molecule type: DNA  
A: Residues: 1-173 <LIN>  
A: Cross-references: EMBL:M13339; NID:g147074; PID:g147075  
R: Lund, B.; Lindberg, F.; Norrback, K.

J. Bacteriol. 170, 1887-1894, 1988  
 Title: Structure and antigenic properties of the tip-  
 Reference number: A27743; PMID:88159520  
 Accession: 13774

```

Length: 22143
Molecule type: DNA
Residues: 1-173 <LUN>
Cross-References: EMBL:M13239; NTD:G147074; NTD:G147075

```

Marklund, B. I.; Tennent, J. M.; Garcia, O. G.; Pol. Microbiol. 6, 2225-2242, 1992  
Title: Horizontal gene transfer of the *Escherichia coli* *cad* and *hns* *hly* genes  
Reference number: S55205

Accession: S25322  
Molecule type: DNA  
Residues: 11173 <MAR>  
Cross-references:

Accession: S25227  
Experimental source: strain J96  
NID: g42290; PID: g42299  
EMBL: X61239;  
References:  
Status: nucleic acid sequence not shown; protein not shown

molecule type: DNA  
 Residues: 1-173 <MA2>  
 Cross-references: EMBL:X61238; NID:q42526; PID:q42527  
 Experimental source: [Stratagene](#)

Note: This protein is one of the minor components of pill. Pill with a defective nucleotide sequence was submitted to the EMBL Data Library, August 1991

A:Gene: pape; prse  
C:Superfamily: pape fibrinri protein  
C:Keywords: fibrinri; pili adhesion  
F:1-24/Domain: signal sequence #status predicted <Sig

Query Match

15.38;

Score 62; DB

15/11/2015

Best Local  
Matches 2

27.08;  
tive

Pred. No. 3.3,  
1; Mismatches

Indels

8; Gaps 3;

QY 1 MKVIV-AEINQEGENKRVFINNVSNIINOMN..... IIII : IIII

DD /9 MAY 11 AM 1968

10

RESULT 15

Species: *Thermotoga maritima*

C:\Accession: 562737

Biochim. Biophys. Acta 1264, 279-283, 1995

A; Reference number: SbZ/3/; MOLD: 50150030

A;Molecule type: DNA  
A;Status: preliminary

A:Cross-references: EMBL:U27841; NID:g8881493; PID:g8881494

Query Match	Score	Pred.	No.
13.16;	35.48;	16;	

Matches	10	conservation	77

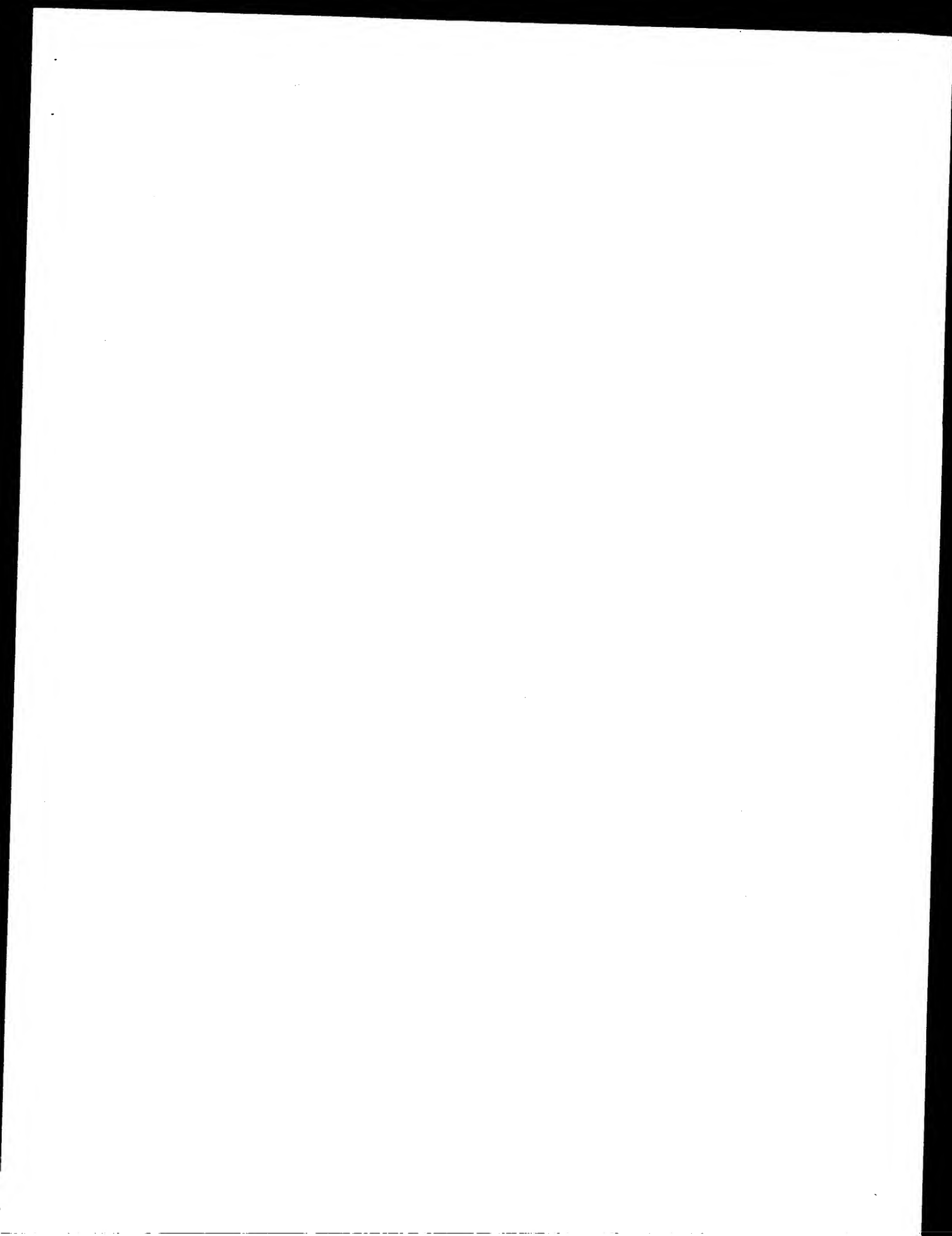
[illegible]

167

```

search completed: September 29, 1999, 13:35:51
job time: 2091 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 29, 1999, 13:14:01 ; Search time 21.98 Seconds

(without alignments)  
102.887 Million cell updates/sec

Title: US-09-142-524-1

Perfect score: 406

Sequence: 1 MVTVAFNQFGPNRRVFYIKR.....IASRVDGLIAYQNPASMK 80

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database: SwissProt\_37.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	24.4	514	1	MPA2_CRYJA
2	71	17.5	374	1	SBP_CRYJA
3	70	17.2	132	1	RS18_HUMAN
4	66.5	16.4	196	1	YMO7_YEAST
5	62.5	15.4	434	1	PEL_LILLO
6	62.5	15.4	1290	1	P19174 human
7	62.5	15.4	145	1	YPI2_AGRTE
8	62	15.3	173	1	PAPF_ECOLI
9	61.3	15.1	633	1	TOP1_THEMA
10	61	15.0	152	1	RS18_DROME
11	60.5	14.9	110	1	Y12K_SMSV4
12	59.5	14.7	842	1	PHS2_RAT
13	59	14.5	862	1	FKH2_YEAST
14	58.5	14.4	1291	1	PIP4_YEAST
15	58.5	14.4	478	1	YSP3_YEAST
16	58	14.3	646	1	HEMA_LADE1
17	58	14.3	656	1	NODD_RHIS3
18	58	14.3	152	1	RS18_ARATH
19	58	14.3	1070	1	YHVA_YEAST
20	57.5	14.2	207	1	PIP4_MOUSE
21	57.5	14.2	1290	1	PIP4_RAT
22	57	14.0	567	1	ODP2_HAEIN
23	57	14.0	511	1	UDPG_DICDI
24	57	14.0	379	1	YLEU_DBOC
25	56.5	13.9	339	1	BMPA_BORBU
26	56.5	13.9	842	1	PHS2_HUMAN
27	56.5	13.9	994	1	PPOL_DROME
28	56.5	13.9	215	1	RAN_BROMA
29	56.5	13.9	1163	1	PROD_PEA
30	56.5	13.8	428	1	AROA_BACSU
31	56	13.8	486	1	LACI_BOTCI
32	56	13.8	833	1	NODD_RHISB
33	56	13.8	843	1	PHS3_HUMAN
34	56	13.8	267	1	RS3_MYCCE
35	55.5	13.7	842	1	PHS2_RABIT
36	55.5	13.7	1305	1	RRLP_AHSV9
37	55.5	13.7	277	1	SET_HUMAN
38	55.5	13.5	185	1	NIR_LEPWC
39	55	13.5	629	1	ODP2_ECOLI
40	55	13.5	470	1	RECA_CHLTR
41	54.5	13.4	470	1	DLDI_BACSU
42	54.5	13.4	413	1	L756_CABEL
43	54.5	13.4	413	1	L756_CABEL

## ALIGNMENTS

44	54.5	13.4	532	1	PROK_AKNSU
45	54	13.3	1296	1	ASAI_ENTFA
					009460 anaerobiosp
					P17953 enterococcu
RESULT 1					
MPA2_CRYJA	STANDARD:	PRT:	514 AA.		
AC P43212;					
DT 01-NOV-1995 (REL. 32, CREATED)					
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)					
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)					
DE POSSIBLE POLYGALACTURONASE PRECURSOR (BC 3.2.1.15) (PG) (PECTINASE)					
DE (MAJOR POLLEN ALLERGEN CRY J 2) (CRY J II).					
OS CRYPTOMERIA JAPONICA (JAPANESE CEDAR).					
OC EUKARYOTA: VIRIDIPHYTES: EMBRYOPHYTES: TRACHEOPHYTES:					
OC EUPHYLLIPHYTES: SPERMATOPHYTES: CONIFEROPSIDA: CONIFERALES:					
OC TAXODIACEAE: CRYPTOMERIA.					
OC [1]					
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.					
RP TISSUE-POLLEN.					
RC MEDLINE: 95010777.					
RX NABHA M., KUROSE M., TORIGOE K., HINO K., TANIGUCHI Y., FUKUDA S.,					
RA USUI M., KURIMOTO M.;					
RA "Molecular cloning of the second major allergen, Cry j II, from					
RT Japanese cedar pollen.";					
RL FEBS LETT. 353:124-128(1994).					
RM [2]					
RP SEQUENCE FROM N.A.					
RC TISSUE-POLLEN.					
RX MEDLINE: 94271166.					
RA KOMIYAMA N., SONE T., SHIMIZU K., MORIKUBO K., KINO K.;					
RT "CDNA cloning and expression of Cry j II the second major allergen of					
RL Japanese cedar pollen.";					
RM BIOCHEM. BIOPHYS. RES. COMMUN. 201:1021-1028(1994).					
RN [3]					
RP SEQUENCE OF 55-64.					
RX MEDLINE: 90342988.					
RA SAKAGUCHI M., INOUE S., TANAI M., ANDO S., USUI M., MATUNASI T.;					
RT "Identification of the second major allergen of Japanese cedar					
RL pollen.";					
RM ALLERGY 45:309-312(1990).					
CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-ALPHA-D-					
CC GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.					
CC -1- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).					
CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES					
CC (POLYGALACTURONASES).					
CC -----					
CC This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC the European Bioinformatics Institute. There are no restrictions on its					
CC use by non-profit institutions as long as its content is in no way					
CC modified and this statement is not removed. Usage by and for commercial					
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>					
CC or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).					
CC -----					
DR EMBL: D37765; G577696; -;					
DR EMBL: D29772; G506858; -;					
DR PROSITE: PS00502; POLYGALACTURONASE; 1.					
DR PFM: PF00295; PG; 1.					
KW AMYLOPLAST; GLYCOPROTEIN; ALLERGEN.					
FT SIGNAL	1	45			POTENTIAL.
FT PROPEP	?	?			POTENTIAL.
FT CHAIN	46	433			POSSIBLE POLYGALACTURONASE.
FT PROPEP	434	514			POTENTIAL.
FT ACT_SITE	278	278			POTABLE.
FT CARBOHYD	460	460			POTENTIAL.
FT CARBOHYD	472	472			POTENTIAL.
FT CONFLICT	5	5			F -> L (IN REF. 2).
FT CONFLICT	12	12			V -> L (IN REF. 2).

```

CC CONFLICT 34 35 DI -> VV (IN REF. 2).
CC CONFLICT 37 37 Q -> K (IN REF. 2).
CC CONFLICT 88 88 K -> N (IN REF. 2).
CC CONFLICT 98 98 N -> S (IN REF. 2).
CC CONFLICT 451 451 K -> E (IN REF. 2).
CC CONFLICT 454 454 G -> R (IN REF. 2).
CC CONFLICT 504 504 M -> I (IN REF. 2).
CC CONFLICT 507 507 R -> C (IN REF. 2).
CC SEQUENCE 514 AA: 56645 MW; 467FE33 CRC32;

Query Match
Best Local Similarity 100.0%; Score 99; DB 1; Length 514;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 31 IDIFASKNFHLOKNTIGTG 49
236 IDIFASKNFHLOKNTIGTG 254

RESULT 2
SBP_CRYUA STANDARD; PRT; 374 AA.
ID SBP_CRYUA
AC P18632;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE SUGI BASIC PROTEIN PRECURSOR (SBP) (MAJOR ALLERGEN CRY J 1) (CRY J 1).
OS CRYPTOMERIA JAPONICA (JAPANESE CEDAR).
OC EUKARYOTA: VIRIDIPHYTES: STREPTOPHYTES: EMBRYOPHYTES: TRACHEOPHYTES:
OC EUPHYLOPHYTES: SPERMATOPHYTES: CONIFEROPSIDA: CONIFERALES;
OC TAXODIACEAE: CRYPTOMERIA.
[1]
RN RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=POLLEN.
RX MEDLINE: 94183234.
RA SONE T., KOMITAMA N., SHIMIZU K., KUSAKABE T., MORIKUBO K.,
RA KINO K.,
RT "Cloning and sequencing of cDNA coding for Cry j 1, a major allergen
RT of Japanese cedar pollen."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 199:619-625(1994).
[2]
RN RN SEQUENCE FROM N.A.
RC TISSUE=POLLEN.
RA NAKAMA M., KUROSE M., TORIGOE K., FUKUDA S., KURIMOTO M.;
RL SUBMITTED (JUL-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
[3]
RN RN SEQUENCE OF 22-41.
RC TISSUE=POLLEN.
RX MEDLINE: 89031257.
RA TANAI M., ANDO S., USUI M., KURIMOTO M., SAKAGUCHI M., INOUE S.,
RA MATSUI T.;
RT "N-terminal amino acid sequence of a major allergen of Japanese cedar
RT pollen (Cry j 1)."
RL FEBS LETT. 239:329-332(1988).
[4]
RN RN CARBOHYDRATES.
RX MEDLINE: 95003748.
RA HIRAKATA A., MATSUMOTO I., KOJIMA K., OGAWA H.;
RT "Antigenicity of the oligosaccharide moiety of the Japanese cedar
RT (Cryptomeria japonica) pollen allergen, Cry j 1."
RL INT. ARCH. ALLERGY IMMUNOL. 105:198-202(1994).
CC -1- P1M: CONTAINS FUCOSE/XILOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.
CC -1- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR
CC POLLEN, THE MOST COMMON POLLEN ALLERGEN IN JAPAN.
CC -1- THE SEQUENCE OF CRY J 1 FORM A IS SHOWN HERE. FORM B DIFFERS IN
CC SIX POSITIONS.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC AMB A I/AMB A II/CRY J I SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D26544; G493632; -.
CC EMBL: D26545; G493636; -.
CC EMBL: D34639; G516726; -.
CC PIR: A44773; A44773.
CC PFAM: PF00544; pec_lyase; 1.
CC ALLERGEN; GLYCOPROTEIN; MULTIGENE FAMILY; SIGNAL.
CC SIGNAL
CC FT SIGNAL 1 21
CC FT CHAIN 22 374 SUGI BASIC PROTEIN.
CC FT VARIANT 12 12 L -> F (IN CRY J 1-B).
CC FT VARIANT 143 143 H -> Y (IN CRY J 1-B).
CC FT VARIANT 202 202 S -> T (IN CRY J 1-B).
CC FT VARIANT 221 221 L -> S (IN CRY J 1-B).
CC FT VARIANT 358 358 Q -> H (IN CRY J 1-B).
CC FT VARIANT 361 361 K -> Q (IN CRY J 1-B).
CC FT CARBOHYD 158 158 POTENTIAL.
CC FT CARBOHYD 191 191 POTENTIAL.
CC FT CARBOHYD 293 293 POTENTIAL.
CC SEQUENCE 374 AA: 40645 MW; DC088629 CRC32;

Query Match
Best Local Similarity 17.5%; Score 71; DB 1; Length 374;
Matches 15; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Db 1 MKYVAFNFGPRFRVIRVSNVH 27
234 MKYVAFNFGPRGCGMPRARGLVH 260

RESULT 3
RS18 HUMAN STANDARD; PRT; 152 AA.
ID RS18 HUMAN
AC P25232;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE 40S RIBOSOMAL PROTEIN S18 (KE-3) (KE3).
GN RPS18.
OS HOMO SAPIENS (HUMAN). RATTUS NORVEGICUS (RAT), MUS MUSCULUS (MOUSE),
OS AND SUS SCROFA (PIG).
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA:
OC PRIMATES: CATARRHINI: HOMINIDAE: HOMO.
[1]
RN RN SEQUENCE FROM N.A.
RC SPECIES=HUMAN; TISSUE=PLACENTA.
RX MEDLINE: 93181276.
RA CHASSIN D., BELLET D., ROMAN A.;
RT "The human homolog of ribosomal protein S18."
RL NUCLEIC ACIDS RES. 21:745-745(1993).
[2]
RN RN SEQUENCE FROM N.A.
RC SPECIES=RAT; STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX MEDLINE: 91337062.
RA CHAN Y.-L., PAZ V., WOOL I.G.;
RT "The primary structure of rat ribosomal protein S18."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 178:1212-1218(1991).
[3]
RN RN SEQUENCE FROM N.A.
RC SPECIES=MOUSE;
RX MEDLINE: 92182530.
RA MACMURRAY A.J., SHIN H.S.;
RT "The murine mhc encodes a mammalian homolog of bacterial ribosomal
RT protein S18."
RL MAMM. GENOME 2:87-95(1992).
[4]
RN RN SEQUENCE FROM N.A.
RC SPECIES=PIG;
RX MEDLINE: 92182530.
RA KIMURA M., KANAKAMI K., SUZUKI H., HAMASIMA N.;

```



Best Local Similarity 28.4%; Pred. No. 4.2;  
Matches 25; Conservative 11; Mismatches 33; Indels 19; Gaps 4;

QY 1 MKVYAFNPGPN-----RRVFKRVSNTIIGRRIDIFASKNFHLQKTTGTGR- 52  
DB 296 MOVYAFNHFGRGLVGRMPCRGFVHYVNDYTHIMYAVGSOH---PILISQGNRY 351  
QY 52 ISLKLTSKIASRRVDGIIAAYONPASW 79  
DB 352 IAPHIEAKEVTKR-----DYAEPAEW 373

## RESULT 6

PIP4\_HUMAN STANDARD; PRT; 1290 AA.

AC P19174;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DE 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE GAMMA 1  
DE (EC 3.1.4.11) (PLC-GAMMA-1) (PHOSPHOLIPASE C-GAMMA-1) (PLC-II)  
DE (PLC-148).  
GN PLCG1 OR PLC1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
AN [1]  
AA SEQUENCE FROM N.A.  
AV 11SSUE-VEIN; AND BRAIN;  
RX BURGESS W.H., DIONNE C.A., KAPLOW J.M., MUDD R., FRIESEL R.,  
PA 71BERSTEIN A., SCHLESINGER J., JAYE M.;  
RT "Characterization and cDNA cloning of phospholipase C-gamma, a major  
RT substrate for heparin-binding growth factor 1 (acidic fibroblast  
RT growth factor)-activated tyrosine kinase.";  
RL MOL. CELL. BIOL. 10:4770-4777(1990).  
RN [2]  
RA BRIDGMAN A.;  
RT SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RX STRUCTURE BY NMR OF SH3 DOMAIN.  
RX MEDLINE; 93208890.  
RA KOHDA D., HATANAKA H., ODAKA M., MANDIYAN V., ULLRICH A.,  
RA SCHLESINGER J., INAGAKI F.;  
RT "Solution structure of the SH3 domain of phospholipase C-gamma.";  
RL CELL 72:953-960(1993)  
CC -1- FUNCTION: PLC-GAMMA IS A MAJOR SUBSTRATE FOR HEPARIN-BINDING  
CC GROWTH FACTOR 1 (ACIDIC FIBROBLAST GROWTH FACTOR)-ACTIVATED  
CC TYROSINE KINASE.  
CC -1- CATALYTIC ACTIVITY: 1-PHOSPHATIDYL-D-MYO-INOSITOL 4,5-BISPHOSPHATE  
CC + H(2)O -> D-MYO-INOSITOL 1,4,5-TRIPHOSPHATE + DIACYLGLYCEROL.  
CC -1- COFACTOR: REQUIRES CALCIUM.  
CC -1- PTM: THE RECEPTOR-MEDIATED ACTIVATION OF PLC-GAMMA 1 AND 2  
CC INVOLVES THEIR PHOSPHORYLATION BY TYROSINE KINASES IN RESPONSE  
CC TO LIGATION OF A VARIETY OF GROWTH FACTOR RECEPTORS AND IMMUNE  
CC SYSTEM RECEPTORS.  
CC -1- THERE ARE AT LEAST SIX FORMS OF PLC ENZYMES.  
CC -1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS  
CC OF PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.  
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS. THE SECOND ONE IS IN TWO  
CC PARTS.  
CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M34667; G190038; -;  
DR EMBL; AL022394; E1312192; -;  
DR PIR; A36466; A36466;  
DR PDB; 2HSP; 31-AUG-94.  
DR PDB; 1HSQ; 31-AUG-94.  
DR MIM; 172420; -;  
DR PROSITE; PS00018; EF\_HAND; 1.  
DR PROSITE; PS50001; SH2; 2.  
DR PROSITE; PS50002; SH3; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 2.  
DR PROSITE; PS50004; C2\_DOMAIN; 2; 1.  
DR PROSITE; PS50007; PIP2C\_X\_DOMAIN; 1.  
DR PROSITE; PS50008; PIP2C\_Y\_DOMAIN; 1.  
DR PFAM; PF00017; SH2; 2.  
DR PFAM; PF00018; SH3; 1.  
DR PFAM; PF00168; SH3; 1.  
DR PFAM; PF00169; PH; 2.  
DR PFAM; PF00387; PI-PLC-Y; 1.  
DR PFAM; PF00388; PI-PLC-X; 1.  
KW HYDROLASE; LIPID DEGRADATION; TRANSUCER; SH2 DOMAIN; SH3 DOMAIN;  
KW CALCIUM-BINDING; PHOSPHORYLATION; 3D-STRUCTURE.  
FT DOMAIN; 27 142  
FT CA\_BIND 165 176  
FT DOMAIN 320 464  
FT DOMAIN 489 523  
FT DOMAIN 550 657  
FT DOMAIN 668 756  
FT DOMAIN 791 851  
FT DOMAIN 895 931  
FT DOMAIN 953 1070  
FT DOMAIN 1075 1177  
FT ACT\_SITE 335 335  
FT ACT\_SITE 380 380  
FT MOD\_RES 771 771  
FT MOD\_RES 783 783  
FT MOD\_RES 1253 1253  
SQ SEQUENCE 1290 AA; 148531 MW; D3DC3E9D CRC32;

## Query Match

15.4%; Score 62.5; DB 1; Length 1290;

Best Local Similarity 28.1%; Pred. No. 14;  
Matches 18; Conservative 9; Mismatches 34; Indels 3; Gaps 1;

QY 5 VAFNPGPNRRVFKRVSNTIIGRRIDIFASKNFHLQ--KNITGTGRISLKTSGKI 61  
DB 883 IAIPEGKNNRLFFVFSISMASVAMSLDVAAADSOEFLDWVKIRREVAQTAADARLTGKI 942  
QY 62 ASRR 65  
DB 943 MERR 946

## RESULT 7

YF12\_AGR6 STANDARD; PRT; 145 AA.

AC Q44433;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 15.8 KD PROTEIN IN PIN2 3'REGION (ORF2).  
OS AGROBACTERIUM TUMEFACIENS.  
OC PLASMID P19A6.  
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;  
OC RHIZOBIACEAE; AGROBACTERIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89213933.  
RA KANDIMOTO R.H., POWELL A.T., AKIYOSHI D.E., REGIER D.A.,  
RA KESTETTER R.A., NESTER E.W., HAWES M.C., GORDON M.P.;  
RT "Nucleotide sequence and analysis of the plant-inducible locus pinf  
RT from Agrobacterium tumefaciens.";  
RL J. BACTERIOL. 171:2506-2512(1989).

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M19352; G1088264; -  
 KW HYPOTHETICAL PROTEIN; PLASMAID.  
 SQ SEQUENCE 145 AA; 15829 MW; 0BF23123 CRC32;  
 Query Match 15.4%; Score 62.5; DB 1; Length 145;  
 Best Local Similarity 33.3%; Pred. No. 1.2;  
 Matches 17; Conservative 11; Mismatches 18; Indels 5; Gaps 2;  
 QY 18 IKRVSNIYH-GRRIDIFASKNFHLQKNTIGTGRISLTKSGIASRRVD 67  
 DB 1 MKRSTLVGVFLATPYVADNIH---TIGTLEIELATLACKPVTYVD 47  
 RESULT 8  
 PAPER ECOLI STANDARD; PRT; 173 AA.  
 AC P08407;  
 DT 01-AUG-1988 (REL. 08, CREATED)  
 DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE BACTERIAL PROTEIN PAPER PRECURSOR.  
 DE PAPER.  
 OS ESCHERICHIA COLI.  
 CC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
 CC ESCHERICHIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J96;  
 RX MEDLINE; 86149403.  
 RA LINDBERG F., LUND B., NORMARK S.;  
 RT "gene products specifying adhesion of uropathogenic Escherichia coli  
 RT are minor components of pili.";  
 RL PROC. NATL. ACADE. SCI. U.S.A. 83:1891-1895(1986).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J96;  
 RX MEDLINE; 88169520.  
 RA LUND B., LINDBERG F., NORMARK S.;  
 RT "Structure and antigenic properties of the tip-located P plus  
 RT proteins of uropathogenic Escherichia coli.";  
 RL J. BACTERIOL. 170:1887-1894(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J96;  
 RX MEDLINE; 93023852.  
 RA MARKLUND B.-I., TENNENT J.M., GARCIA E., HAMERS A., BAGA M.,  
 RA LINDBERG F., GASTRA W., NORMARK S.;  
 RT "Horizontal gene transfer of the Escherichia coli pap and prs pili  
 RT operators as a mechanism for the development of tissue-specific  
 RT adhesive properties.";  
 RL MOL. MICROBIOL. 6:2225-2242(1992).  
 CC -1- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING  
 CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5  
 CC MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO  
 CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.  
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE MINOR COMPONENTS OF PILI.  
 CC PILI WITH A DEFECTIVE PAPER GENE WILL HAVE LOW ADHESIVE CAPACITY  
 CC OR NONE; HOWEVER, THE BINDING PROPERTY OF THE WHOLE CELL WILL NOT  
 CC BE AFFECTED.  
 CC -1- SUBCELLULAR LOCATION: SECRETED; TIP OF THE PILI.  
 CC -1- DISEASE: STRAINS OF E.COLI THAT CAUSE INFECTION OF THE HUMAN  
 CC URINARY TRACT PRODUCE PAP-PILI WHICH ARE HAIR-LIKE APPENDAGES  
 CC CONSISTING OF ABOUT 1000 HELICALLY ARRANGED SUBUNITS OF THE

-----  
 CC PROTEIN PAPA. THESE PILI MEDIATE BINDING TO DIGALACTOSIDE-  
 CC CONTAINING GLYCOLIPIDS PRESENT ON THE EPITHELIAL CELLS WHICH  
 CC LINE THE URINARY TRACT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X61238; G42527; -  
 DR EMBL: M13239; G147075; -  
 DR EMBL: M20146; G147090; -  
 DR EMBL: X61239; G42299; -  
 DR PIR: A25134; YOECPD.  
 DR PFAM: PF00419; Fimbrin; 1.  
 KW FIMBRIA; CELL ADHESION; SIGNAL.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 173 FIMBRIAL PROTEIN PAPER.  
 SQ SEQUENCE 173 AA; 18569 MW; DA6F5994 CRC32;  
 Query Match 15.3%; Score 62; DB 1; Length 173;  
 Best Local Similarity 27.0%; Pred. No. 1.7;  
 Matches 20; Conservative 11; Mismatches 35; Indels 8; Gaps 3;  
 QY 1 MKVTV-AFNGQPNRVRVIRKVSNIYHGRRIDIFASKNFHLQKNTIGTGRISLTKSG 59  
 DB 79 MKVITATNTY--NNAILVQNTSNDGLVLYNS-----NMGNTGATLTGTPPTG 131  
 QY 60 KIASRRVDCITLAAV 73  
 DB 132 KITGNADKTIISLH 145  
 RESULT 9  
 TOP1\_THEMA STANDARD; PRT; 633 AA.  
 AC P46799;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE DNA TOPOISOMERASE I (EC 5.99.1.2) (OMEGA-PROTEIN) (RELAXING ENZYME)  
 DE (UNWINDING ENZYME) (SWIVELASE).  
 GN TOPA.  
 OS THERMOTOGA MARITIMA.  
 CC BACTERIA; THERMOTOGALES; THERMOTOGA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MSB8 / DSM 3109;  
 RA LA TOUR C., KALTOUM H., PORTEMER C., CONFALONIERI F., HUBERT R.,  
 RA DUGUET M.;  
 RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE  
 CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.  
 CC -1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED  
 CC DNA, FOLLOWED BY PASSAGE AND REJOINING.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA BACKBONE BOND, IT  
 CC SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN WHICH A TYROSIL OXYGEN  
 CC IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS AT ONE END OF THE  
 CC ENZYME-SEVERED DNA STRAND.  
 CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL: U27841; G881494; -  
 DR PROSITE; PS00396; TOPOISOMERASE\_I\_PROK; 1.  
 DR PFAM: PF01131; Topoisom\_bac; 1.  
 DR HSSP: P06612; 1EC1.  
 KW ISOMERASE; TOPOISOMERASE; DNA-BINDING.  
 FT ACT\_SITE 288  
 SQ SEQUENCE 633 AA; 72694 MW; 9CDD9F68 CRC32;

Query Match 15.1%; Score 61.5; DB 1; Length 633;  
 Best Local Similarity 25.4%; Pred. No. 8.5;  
 Matches 16; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 18 IKRVSNIYHGRIDIFASKNFHLQKNTIGTGRRISLKTSGKASRYDGIIAAYONPA 77  
 DB 96 IARVTNLTGRKNRI-VESEITPRVIREAVKPREIDMKVRAQLARIIDRTVGSLSPIV 154  
 QY 78 SWK 80  
 DB 155 LWR 157

RESULT 10  
 RS18\_DROME STANDARD; PRT; 152 AA.

AC P41094;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
 DE 40S RIBOSOMAL PROTEIN S18.  
 GN RPS18.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOPORPHA; EPHYROIDEA;  
 OC DROSOPHILINAE; DROSOPHILA.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON-S;  
 RA MEDLINE: 94215909.  
 RX GARWOOD J., LEPEZANT J.A.;

RT "The Drosophila melanogaster homolog of ribosomal protein S18.";  
 RL GENE 141:231-235(1994).

CC - FUNCTION: INVOLVED IN THE BINDING OF PMET-TRNA AND, HENCE, IN THE INITIATION OF TRANSLATION (BY SIMILARITY).  
 CC - SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L22959; G349256; -  
 DR FLYBASE; FBgn0010411; RPS18.  
 DR PROSITE; PS00646; RIBOSOMAL\_S13; 1.  
 DR PFAM: PF00416; S13; 1.  
 KW RIBOSOMAL PROTEIN  
 SQ SEQUENCE 152 AA; 17611 MW; 72CA0DCB CRC32;

Query Match 15.0%; Score 61; DB 1; Length 152;  
 Best Local Similarity 25.6%; Pred. No. 2;  
 Matches 21; Conservative 10; Mismatches 29; Indels 22; Gaps 3;

QY 12 PNRNRVIRKVSNIYHGR--IDIFASKNFHLQKNTIGTGR-----ISLKTLS 58  
 DB 6 PEKFOHLIRMTNIDGKRKRVGIAMTAIK-----GVGRYSNIYVKADVDLTAKA 56  
 RX 59 GKIASRRVDSIIAAYONPASWK 80

DB 57 GECTEEVDKVTIISNPLOK 78

RESULT 11  
 Y12K\_SMSV4 STANDARD; PRT; 110 AA.

AC P36289;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE POSSIBLE 12 KD NUCLEIC ACID-BINDING PROTEIN.  
 OS SAN MIGUEL SEA LION VIRUS (SEROTYPE 4) (SMSV 4)  
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES; NO DNA STAGE; CALICIVIRIDAE;  
 OC CALICIVIRUS.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92410750.

RA NELL J.D.;

RT "Nucleotide sequence of the capsid protein gene of two serotypes of San Miguel sea lion virus: identification of conserved and non-conserved amino acid sequences among calicivirus capsid proteins.";  
 RL VIRUS RES. 24:211-222(1992).

CC - SIMILARITY: TO FELINE CALICIVIRUS 12 KD PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M87482; G334889; -  
 DR PIR: D48562; D48562.  
 SQ SEQUENCE 110 AA; 12566 MW; 24096EF1 CRC32;

Query Match 14.9%; Score 60.5; DB 1; Length 110;  
 Best Local Similarity 32.4%; Pred. No. 1.6;  
 Matches 11; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 17 FIKRVSNIYHGRIDIFASKNFHLQKNTIGTGR 50  
 DB 10 FLNSVANAYVEGKLDL-ASKGLQLKSRALDTER 42

RESULT 12

PHS2\_RAT STANDARD; PRT; 842 AA.

AC P09812;  
 DT 01-MAR-1989 (REL. 10, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE GLYCOGEN PHOSPHORYLASE, MUSCLE FORM (BC 2.4.1.1) (MYOPHOSPHORYLASE).  
 GN PYGM.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.

RN (1)  
 RP SEQUENCE OF 2-842 FROM N.A.  
 RA HUDSON J.W., HEFFERON K.L., CREERAR M.M.;  
 RL SUBMITTED (OCT-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN (2)  
 RP SEQUENCE OF 566-762 FROM N.A.  
 RX MEDLINE: 86248098

RA OSAMA S., CHIU R.H., MCDONOUGH A., MILLER T.B. JR., JOHNSON G.L.;

RT "Isolation of partial cDNAs for rat liver and muscle glycogen phosphorylase isozymes.";  
 RL FEBS LETT. 202:282-288(1986).  
 RN (3)

RP SEQUENCE OF 763-842 FROM N.A.  
 RX MEDLINE: 86030264.

RA HWANG P.K., SEE Y.P., VINCENTINI A.M., POWERS M.A., FLETTERICK R.J.,  
 RA CREAR M.M.:  
 RT "Comparative sequence analysis of rat, rabbit, and human muscle  
 glyco-gen phosphorylase cDNAs."  
 CC FOR. J. BIOCHEM. 152:267-274(1985).  
 CC -1- FUNCTION: PHOSPHORYLASE IS AN IMPORTANT ALLOSTERIC ENZYME IN  
 CARBOHYDRATE METABOLISM. ENZYMES FROM DIFFERENT SOURCES DIFFER IN  
 THEIR REGULATORY MECHANISMS AND IN THEIR NATURAL SUBSTRATES.  
 CC HOWEVER, ALL KNOWN PHOSPHORYLASES SHARE CATALYTIC AND STRUCTURAL  
 PROPERTIES.  
 CC -1- CATALYTIC ACTIVITY: (1,4-ALPHA-D-GLUCOSYL)(N) + ORTHOPHOSPHATE =  
 (1,4-ALPHA-D-GLUCOSYL)(N-1) + ALPHA-D-GLUCOSE 1-PHOSPHATE.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- SUBUNIT: HOMODIMER. DIMERS ASSOCIATE INTO A TETRAMER TO FORM THE  
 ENZYMATICALLY ACTIVE PHOSPHORYLASE A.  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOGEN PHOSPHORYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L10669; G204423; -  
 DR EMBL: X03032; G818028; -  
 DR PIR: B24200; B24200.  
 DR PIR: C23093; C23093.  
 DR PROSITE: PS00102; PHOSPHORYLASE; 1.  
 DR PFAM: PF00343; phosphorylase; 1.  
 DR HSP: P00489; 1Pys.  
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; CARBOHYDRATE METABOLISM;  
 KW GLYCOGEN METABOLISM; ALLOSTERIC ENZYME; PYRIDOXAL PHOSPHATE;  
 KW PHOSPHORYLATION.  
 FT MOD\_RES 15 15  
 FT BINDING 76 76  
 FT SITE 109 109  
 FT SITE 143 143  
 FT ACT\_SITE 156 156  
 FT BINDING 661 661  
 FT CONFLICT 640 641  
 FT CONFLICT 724 724  
 FT CONFLICT 766 766  
 FT SEQUENCE 842 AA; 97273 MW; 32741482 CRC32;  
 SO  
 Query Match 14.7%; Score 59.5; DB 1; Length 842;  
 Best Local Similarity 29.7%; Pred. No. 20;  
 Matches 22; Conservative 9; Mismatches 30; Indels 13; Gaps 3;  
 QY 10 FGNRRVFIKRVSNVITIGRRIDIFA-----SKNHLQKNTIGTGRRLSLKLTSG 59  
 DB 751 FSPKQDLEFDIYVNMVHHDFKVFADYEYIKQDKVSELYKNPREMTVMYIRNIATSG 810  
 QY 60 KIASRRVDGIITAY 73  
 DB 811 KFS-----DRIIAY 821  
 AC P41813;  
 AC P41813; STANDARD; PRT; 862 AA.  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE FORK HEAD PROTEIN HOMOLOG 2.

GN FK22 OR YNL068C OR N2403 OR YNL2403C.  
 OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMETES; SACCAROMYCETALES;  
 OC SACCAROMYCETACEAE; SACCAROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C.  
 RC STRAIN-S288C.  
 RA ZHU G., DAVIS T.N.;  
 RL SUBMITTED (XXX-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / FY1679;  
 RX MEDLINE: 96267764.  
 RA POEHLMANN R., PHILIPSEN P.;  
 RT "Sequencing a cosmid clone of *Saccharomyces cerevisiae* chromosome XIV  
 reveals 12 new open reading frames (ORFs) and an ancient duplication  
 of six ORFs."  
 RT YEAST 12:391-402(1996).  
 RL [3]  
 RP SEQUENCE OF 1.440 FROM N.A.  
 RC STRAIN-S288C / FY1676;  
 RX MEDLINE: 96021608.  
 RA BERGERZ P., DOIGNON F., CROUZET M.;  
 RT "The sequence of a 44 420 bp fragment located on the left arm of  
 chromosome XIV from *Saccharomyces cerevisiae*."  
 RL YEAST 11:967-974(1995).  
 RN [4]  
 RP ERRATUM.  
 RX MEDLINE: 97060022.  
 RA BERGERZ P., DOIGNON F., CROUZET M.;  
 RL YEAST 12:297-297(1996).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L38850; G623608; -  
 DR EMBL: X86470; G791119; -  
 DR EMBL: Z71343; E328836; -  
 DR EMBL: Z71344; E239903; -  
 DR EMBL: U12141; G1098481; -  
 DR SGD: L0002608; FK22.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50006; FHA\_DOMAIN; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 DR PFAM: PF00250; Fork head; 1.  
 DR PFAM: PF00498; FHA; 1.  
 KW DNA-BINDING; NUCLEAR PROTEIN.  
 FT DOMAIN 83 152  
 FT DNA\_BIND 339 430  
 FT SEQUENCE 862 AA; 94374 MW; 83DA8394 CRC32;  
 SO  
 Query Match 14.5%; Score 59; DB 1; Length 862;  
 Best Local Similarity 30.6%; Pred. No. 24;  
 Matches 22; Conservative 11; Mismatches 31; Indels 8; Gaps 4;  
 QY 11 GPNRRVFIKRVSNVITIGRRIDIFASKNFKLQKNTIG--TGRRLSLKLTSGKIASRRVDG 68  
 DB 71 GPNRTYYKLEYSI--GRNTD--PLNSALQENSDGVKNSYRVNIDLGPAKVSRRK-RA 124  
 QY 69 IIAAYQNPASWK 80  
 DB 125 IIKYNNMGWE 136



RP SEQUENCE FROM N.A.  
 RA MASON O.B., WONG P.A., BARR P.J.:  
 RL SUBMITTED (XXX-1991) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA MEDLINE; 97051599.  
 RA STERKY F., HOLMBERG A., PETERSSON B., UHLEN M.:  
 RT "The sequence of a 30 kb fragment on the left arm of chromosome XV  
 from *Saccharomyces cerevisiae* reveals 15 open reading frames, five of  
 which correspond to previously identified genes."  
 RL YEAST 12:1091-1095(1996).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
 SUBTILASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC in the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

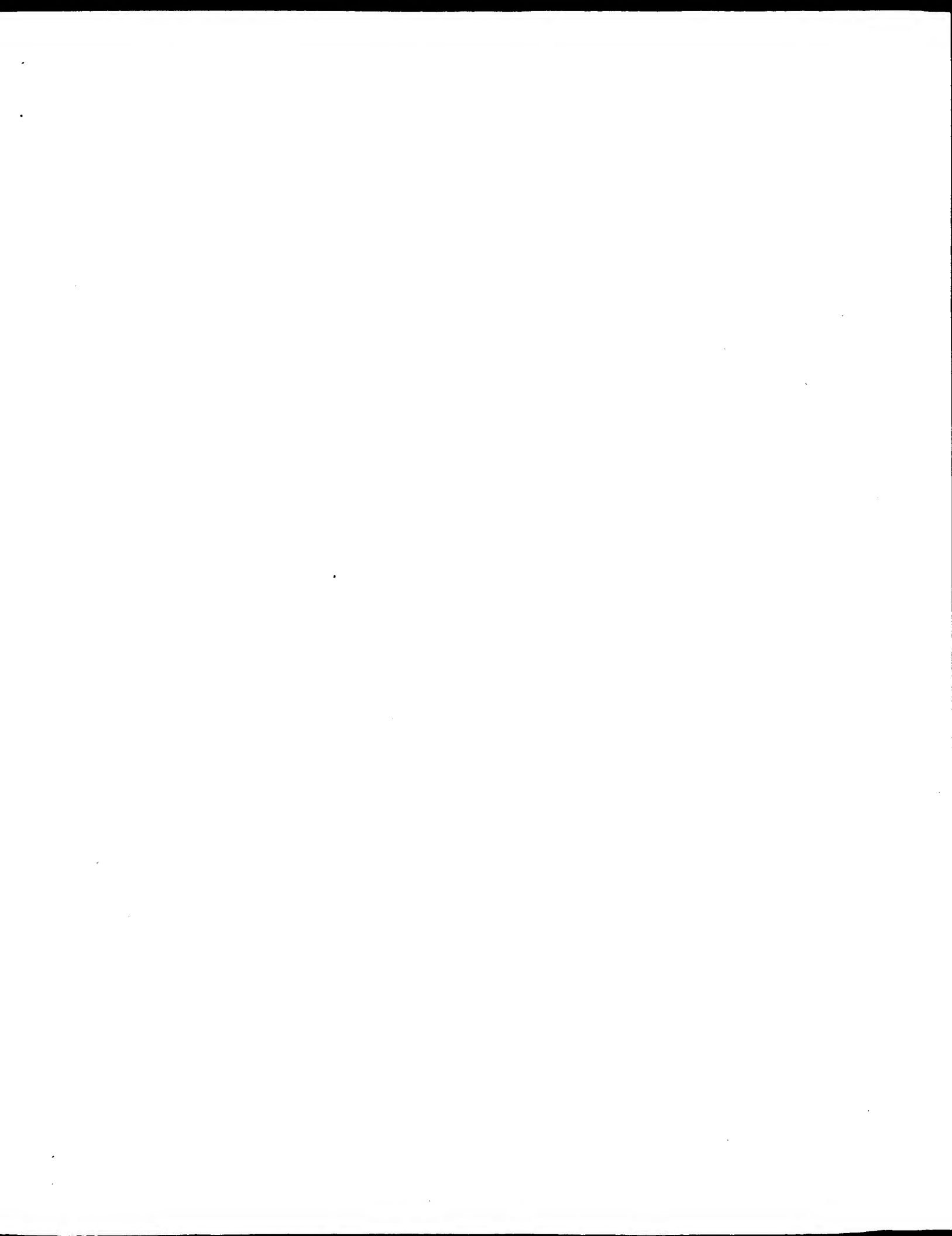
CC  
 DR EMBL; M7197; G173237; -;  
 DR EMBL; U43491; G1150998; -;  
 DR SCD; Z74911; E252316; -;  
 DR SGD; L0002552; YSP3;  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR PFAM; PF00082; subtilase; 1.  
 DR HSSP; P06873; 3PRK.  
 DR KW HYDROLASE; SERINE PROTEASE; SIGNAL.

FT SIGNAL 1 ?  
 FT CHAIN 1 ?  
 FT ACT\_SITE 213 478 SUBTILASIN-LIKE PROTEASE III.  
 FT ACT\_SITE 245 245 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 407 407 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CONFLICT 22 22 D -> E (IN REF. 1).  
 FT CONFLICT 105 105 N -> D (IN REF. 1).  
 FT CONFLICT 189 205 MISSING (IN REF. 1).  
 FT CONFLICT 289 289 Y -> L (IN REF. 1).  
 SO SEQUENCE 478 AA; 52089 MW; 140C57D2 CRC32;

Query Match 14.4% Score 58.5; DB 1; Length 478;  
 Best Local Similarity 29.6% Pred. No. 14;  
 Matches 21; Conservative 14; Mismatches 27; Indels 9; Gaps 4;

QY 3 VTVAFNQFGPNRRVFIKRVSVNIIHGRIRIDFASKNFHLQKNTIGTGRTSLKLTSGKIA 62  
 Db 359 ITVGASTSFSDRAFF---SN---WGTCVDVFA-PGAINMSTYIGS-RNATLSLSTGTSMA 409  
 QY 63 SRRVDGIIAAY 73  
 Db 410 SPHVAGILSYF 420

Search completed: September 29, 1999, 13:37:16  
 Job time: 1395 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 29, 1999, 13:02:00 ; Search time 45.11 Seconds

(without alignments)  
109.144 Million cell updates/sec

Title: US-09-142-524-1

Perfect score: 406

Sequence: 1 MKVTVAFNQFGPNRRVFIKR.....IASRRVDGIIAAYQNPAW 80

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTOZOA:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	18.2	375	10 Q96385	Q96385 chameacypar
2	73	18.0	367	10 Q92NU7	Q92NU7 Juniperus a
3	72	17.7	324	5 Q46197	Q46197 dirosophila
4	68	16.7	394	10 Q65456	Q65456 arabidopsis
5	66	16.3	350	2 P73865	P73865 synechocyst
6	66	16.3	394	10 Q65457	Q65457 arabidopsis
7	65	16.0	390	10 Q65388	Q65388 arabidopsis
8	65	16.0	498	12 Q88525	Q88525 turkey hep
9	64	15.8	222	8 Q21280	Q21280 reclinomona
10	63.5	15.6	533	5 Q01617	Q01617 caenorhabdi
11	61.5	15.1	102	2 Q66004	Q66004 mycoplasma
12	61	15.0	173	2 Q47446	Q47446 escherichia
13	60	14.8	726	4 Q15152	Q15152 homo sapien
14	60	14.8	725	4 Q13835	Q13835 homo sapien
15	60	14.8	275	5 Q44584	Q44584 caenorhabdi
16	60	14.8	587	5 Q43956	Q43956 tetrahymena
17	60	14.8	154	5 Q18240	Q18240 caenorhabdi
18	60	14.8	295	6 Q28875	Q28875 bos taurus
19	60	14.8	727	6 Q28161	Q28161 bos taurus
20	60	14.8	627	10 Q64766	Q64766 arabidopsis
21	60	14.8	257	10 Q92P37	Q92P37 nicotiana t
22	59.5	14.7	254	1 Q30306	Q30306 archaeglob
23	59.5	14.7	269	10 Q48980	Q48980 hordeum vul
24	59	14.5	152	3 Q94754	Q94754 schizosacch
25	58.5	14.4	380	11 P97877	P97877 rattus norv
26	58	14.3	729	1 P95924	P95924 sulfolobus
27	58	14.3	1396	2 Q05450	Q05450 mycobacteri
28	58	14.3	578	2 Q92K18	Q92K18 helicobacte
29	58	14.3	661	5 Q21394	Q21394 caenorhabdi

30	58	14.3	728	11 P97350	P97350 mus musculu
31	57.5	14.2	103	2 Q51514	Q51514 borrelia bu
32	57.5	14.2	314	2 Q92G17	Q92G17 borrelia
33	57.5	14.2	796	5 Q17532	Q17532 caenorhabdi
34	57.5	14.2	834	5 Q19783	Q19783 caenorhabdi
35	57	14.0	616	1 Q51714	Q51714 pyrococcus
36	57	14.0	314	2 Q92ZAN5	Q92ZAN5 comamonas s
37	57	14.0	593	3 Q60102	Q60102 schizosacch
38	57	14.0	656	9 Q38355	Q38355 lactococcus
39	56.5	13.9	254	2 P74730	P74730 synechocyst
40	56.5	13.9	339	2 Q87960	Q87960 borrelia bu
41	56.5	13.9	538	5 Q16342	Q16342 caenorhabdi
42	56.5	13.9	563	5 Q45758	Q45758 caenorhabdi
43	56.5	13.9	431	10 Q23017	Q23017 arabidopsis
44	56.5	13.9	108	12 Q09783	Q09783 san miguel
45	56	13.8	492	4 Q14926	Q14926 homo sapien

## ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	375 AA.
ID Q96385			
AC Q96385;			
DT 01-FEB-1997 (TREMBLrel. 02, Created)			
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)			
DE CHAO1 PRECURSOR.			
OS Chameacyparis obtusa.			
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;			
OC Cupressaceae; Chameacyparis.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE-POLLEN.			
RX MEDLINE; 96265194.			
RA SUZUKI M., KOMIYAMA N., ITOH M., ITOH H., SONE T., KUNO K., TAKAGI I.,			
RA OHTA N.;			
RT "Purification, characterization and molecular cloning of Cha o 1, a			
RT major allergen of Chameacyparis obtusa (Japanese cypress) pollen.";			
RL Mol Immunol. 33:451-460(1996).			
DR EMBL; D45404; BA08246.1; -			
DR MENDEL; 7626; CHAOB; 1088; 1.			
DR PFAM; PF00544; pec_lyase; 1.			
DR SIGNAL.			
FT SIGNAL.			
KW CHAIN			
FT CHAIN			
SO SEQUENCE			
375 AA; 40258 MW; A0981492 CRC32;			
Query Match	18.2%;	Score 74;	DB 10;
Best Local Similarity	59.3%;	Pred. No. 0.51;	Length 375;
Matches	16;	Conservative	2;
Mismatches	9;	Indels	0;
Gaps	0;		
QY 1 MKVTVAFNQFGPNRRVFIKRVSNVH 27			
DB 234 MKVTVAFNQFGPNRQRMFRARYGLH 260			
RESULT 2			
ID Q92NU7			
AC Q92NU7;			
DT 01-MAY-1999 (TREMBLrel. 10, Created)			
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)			
DE POLLEN MAJOR ALLERGEN 1-2.			
OS Juniperus ashei (Ozark white cedar).			
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;			
OC Cupressaceae; Juniperus.			
RN [1]			

RP SEQUENCE FROM N.A.  
RA MIDOHO-HORIUTI T.M., GOLDBLUM R.M., KUROSKY A., WOOD T.G.,  
RA BROOKS E.G.;  
RT "Molecular cloning of mountain cedar (Juniperus ashei) pollen major  
RT allergen, Jun a 1,"  
DR EMBL: A0106663; AAD03609.1;  
DR EMBL: A0106662; AAD03608.1;  
SQ SEQUENCE 367 AA; 39824 MM; 4C2DB630 CRC32;

Query Match 18.0%; Score 73; DB 10; Length 367;

Best Local Similarity 55.6%; Pred. No. 0.65;

Matches 15; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MKVTVAENQFGPNRRVFIKRVSNVTH 27  
DB 234 MKVTVAENQFGPNRRVFIKRVSNVTH 260

RESULT 3  
ID 046197 PRELIMINARY; PRT; 234 AA.  
AC 046197;  
DT 01-JUN-1998 (TREMblrel. 06, Created)  
DT 01-JUN-1998 (TREMblrel. 06, last sequence update)  
DT 01-MAY-1999 (TREMblrel. 10, last annotation update)  
DE ACCESSORY GLAND PROTEIN ACP29AB.  
GN ACP29AB.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;  
OC Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CANTON S; TISSUE-ACCESSORY GLAND;  
RA WOLFEY M.F., HARADA H.A., BERTRAM M.J., STELICK T.J., KRAUS K.W.,  
RA KALB J.M., LUNG Y.O., NEUBAU D.M., PARK M., TRAM U.;  
RL Insect Biochem. Mol. Biol. 27:0-0(1997).  
DR EMBL: 085758; AAB96382.1;  
DR FLXBASE: FB9001583; ACP29AB.  
DR PFAM: PF00059; lectin\_c1.  
SQ SEQUENCE 234 AA; 27173 MM; F674589C CRC32;

Query Match 17.7%; Score 72; DB 5; Length 234;

Best Local Similarity 25.9%; Pred. No. 0.49;

Matches 21; Conservative 19; Mismatches 31; Indels 10; Gaps 3;

QY 9 OFGPNRRVFIKRVSNV---IIHGRIDIFASKNFHLOKNTIGT--GRRISIKLTSGRIA 62  
DB 88 QLOPLKTIHRHASNKSNKMKRFEVSGRHHIENKIMQTFEAYVTCRKNGHILA 147  
OY 63 S---RRVDGIIAAYONPASW 79  
DB 148 NIDKEKELDGILLALAPNNSTW 168

RESULT 4  
ID 065456 PRELIMINARY; PRT; 394 AA.  
AC 065456;  
DT 01-AUG-1998 (TREMblrel. 07, Created)  
DT 01-AUG-1998 (TREMblrel. 07, last sequence update)  
DT 01-MAY-1999 (TREMblrel. 10, last annotation update)  
DE PEPTIDE LIASE LIKE PROTEIN.  
GN FLN20.180.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]

RP SEQUENCE FROM N.A.  
RA BEVAN M., WEDLER H., WAMBUIT R., BANCROFT I., MEWES H.W., MAYER K.,  
RA SCHUELLER C.;  
RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU ARABIDOPSIS SEQUENCING PROJECT;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL022140; CAA18111.1;  
DR PFAM: PF00544; pec\_lyase; 1.  
KM lyase.  
SQ SEQUENCE 394 AA; 43476 MM; 95399178 CRC32;

Query Match 16.7%; Score 68; DB 10; Length 394;

Best Local Similarity 28.6%; Pred. No. 2.8;

Matches 24; Conservative 7; Mismatches 27; Indels 26; Gaps 3;

QY 1 MKVTVAENQFGPNRRVFIKRVSNVTH--GRRID-----IFASKNFHLOKN 44  
DB 256 MKVTVAENQFGPNRRVFIKRVSNVTH--GRRID-----IFASKNFHLOKN 44

QY 45 TIGTGRISIKLTSGKIASRRVDG 68  
DB 316 -----KSNKEVTKREYKG 329

RESULT 5  
ID P73865 PRELIMINARY; PRT; 350 AA.  
AC P73865;  
DT 01-FEB-1997 (TREMblrel. 02, Created)  
DT 01-FEB-1997 (TREMblrel. 02, last sequence update)  
DT 01-JAN-1999 (TREMblrel. 09, last annotation update)  
DE SENSOR TRANSDUCTION HISTIDINE KINASE.  
OS Synecocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;  
RA TABATA S.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;  
RX MEDLINE: 97061201.  
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
RA MIYATA N., HIROGAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
RA TABATA S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL: D90910; BAA17927.1;  
DR PFAM: PF00512; signal; 1.  
SQ SEQUENCE 350 AA; 39445 MM; EEPF2A0B CRC32;

Query Match 16.3%; Score 66; DB 2; Length 350;

Best Local Similarity 39.5%; Pred. No. 4.2;

Matches 17; Conservative 7; Mismatches 15; Indels 4; Gaps 2;

QY 14 RRVFIKRVSNVTH--GRRIDI-FASKNFHLOKNTIGRRT 52  
DB 246 QRVFNILITAINHSPGRKRVETISLTKNNHFOVQIVDEGRGI 288

RESULT 6  
ID 065457 PRELIMINARY; PRT; 394 AA.  
AC 065457;

```

Query Match 16.3% Score 66; DB 10; Length 394;
      58; Local Similarity 28.6%; Pred No. 4.8;
      2; Conservative 7; Mismatches 27; Indels 26; Gaps 3
QY 1 MKTVAFNEHNOGPPRRRFVFKRVSWIHH--GRRID-----IFASKNFHLQKN 44
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45 MKTVAFNEHNGPELVPRMRPRVRGTAHAANNKRYDKIMVYAIGSADPTIFSEGNVFIASD 315
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 45 TIGTGRISLIKLTSGKIASRRVDG 68
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 -----KSYSKVEVTKRREYVG 329
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 7	
ID	065388
AC	065388;
DT	01-AUG-1998 (TREMblrel. 07, Created)
DT	01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT	01-MAY-1999 (TREMblrel. 10, Last annotation update)
DE	F12F1.22 PROTEIN.
GN	F12F1.22.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC	core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC	Arabidopsis.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. COLUMBIA:
RA	LYSOSAKA V.S., OSBORNE B.I., SCHWARTZ J.R., TORIIMI M., YU G.,
RA	KWAN A., OI O., LIO S., BOEHLER E., CONWAY A.B., CONWAY A.R.,
RA	DEMAR K., FENG J., KIM C., KURTZ D., PALM C.J., LI Y., SHINN P.,
RA	SUN H., DAVIS R.W., ECKER J.R., FEDERSPIEL N.A., THEOLOGIS A.,
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBD databases.
DR	EMBL; ACC00213; AAC17625.1; -
DR	PFAM; PF00544; pec_lyase; 1.
SO	SEQUENCE 390 AA; 43354 MW; E0F322FB4 CRC32;

```
Query Match      16.0% Score 65; DB 10; Length 390;  
Best Local Similarity 25.0%; Pred. No. 6.2;  
Matches    25; Conservative   1; Mismatches    29; Indels    28; Gaps    5;
```

QY

```
1 MKVTAAFNQGNRRRVFKRSNYIIH-----GRID--TFASKNFHDKN 44  
|:::||||| |||| | ||| ::||  
Db 246 MGVTLAFNHFGGLERPRVRBRYAVANNRREKMTAIGSADPTFESGNE---- 302
```

RESULT	8		
088525			
ID	088525	PRELIMINARY;	PRT; 498 AA.
AC	088525:		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-NOV-1998	(TREMBLrel. 08, Last annotation update)	
DE	GLYCOPROTEIN HOMOLOGUE E.		
OS	Turkey herpesvirus.		
OC	Viruses; dsDNA, "viruses, no RNA stage; Herpesviridae.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE; 9401499.		
RX	ZELNIK V., DAREIL R., AUDONNET J., SMITH G.D., RIVIERE M.,		
RA	PASTOREK J., ROSS N.L.J.;		
RA	"The complete sequence and gene organization of the short unique		
RT	region of herpesvirus of turkeys."		
RL	J. Gen. Virol. 74:2151-2162(1993).;		
DR	EMBL; X68653; CAA48619.1;		
SO	SEQUENCE 498 AA; 55059 MW; 0E20E22F CRC32;		

```

Query Match 16.0%; Score 65; DB 12; Length 498;
Best Local Similarity 31.6%; Pred NO. 8.4.
Matches 18; Conservative 10; Mismatches 25; Indels 4; Gaps 2

QY 18 I R K V S N I I I H G R R I D I F A S K N F H L O K N T - I T G R R I S I K I T S G I A S R R V D G I I A A Y 73
Db 92 I L R V D F V N S S G L D I A S S C E H I P N I I D T G K E L L E F N - - A T S Q I A G V Y T R Y 145

```

RESULT	9	
021280		
ID	021280	PRELIMINARY;
AC	021280;	PRT; 222 AA.
DT	01-JAN-1998	(TREMBLrel. 05, Created)
DD	01-JAN-1998	(TREMBLrel. 05, Last sequence update)
DT	01-NOV-1998	(TREMBLrel. 08, Last annotation update)
DE	ATP-BINDING SUBUNIT OF ABC TRANSPORTER FOR CYTOCHROME C1.	
GN	YEW.	
OS	Reclinomonas americana.	
OG	Mitochondrion.	
OC	Eukaryota; Reclinomonas.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-ATCC50394;	
RA	MEDLINE; 97311393.	
RX	LANG B.F., BURGER G., O'KELLY C.J., CEDERGREEN R., GOLDING G.B.,	
RA	LEMEUX C., SANKOFF D., TURMEL M., GRAY M.W.;	
RT	"An ancestral mitochondrial DNA resembling a eubacterial genome in	
RL	miniature.";	
RM	Nature 387:493-497(1997).	
RP	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-ATCC50394;	
RA	LANG B.F., BURGER G.;	
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF007261; AAD1907.1; -	
DR	PFAM; PF00005; ABC.ttran.1.	
KW	Atp-binding; Mitochondrion.	
SEQUENCE	222 AA; 25502 MW; D81B02D7 CRC32;	

Query Match 15.8%; Score 64; DB 8; Length 222;  
 Best Local Similarity 30.8%; Pred. No. 4.1.  
 Matches 20; Conservative 14; Mismatches 19; Indels 12; Gaps 3;  
 18 IKRVSNIH---GRID--IFASINFLQKNT-----GTCRRISLITGSKIASRR 65

DB 3 LKIQWLTHTNTGIRSKNTIFONTNFSLEKGSLEPIIOGNSGKTYTLKITSGLPFSQ 62  
 QY 66 VDGII 70  
 DB 63 GDIII 67

RESULT 10  
 ID 001617 PRELIMINARY; PRT: 533 AA.

AC 001617;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
 DE HYPOTHETICAL 60.7 KD PROTEIN T19H12.1 IN CHROMOSOME V PRECURSOR.  
 GN T19H12.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA DAVIDSON S.;  
 RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO FAMILY UPF.  
 DR EMBL: U97009; AAC69026.1;  
 DR WORMPEP: T19H12.1; CEL3752.  
 DR PFAM: PF00201; UDPGT; 1.  
 DR HYPOTHETICAL protein; Signal; Transmembrane.  
 FI SIGNAL 18  
 FI CHAIN 19 533  
 FT TRANSMEM 152 172  
 FT TRANSMEM 260 280  
 FT TRANSMEM 495 515  
 FT TRANSMEM 533 AA; 60726 MW; 008AAFTD CRC32;  
 SQ SEQUENCE

Query Match 15.6%; Score 63.5; DB 5; Length 533;  
 Best Local Similarity 28.2%; Pred. No. 14;  
 Matches 19; Conservative 14; Mismatches 17; Indels 15; Gaps 4;  
 QY 10 FGPNRRVEIKRVSNI-HGRRIDIFASKNFHLQNTIGTGRISLTKTSKASRVVG 68  
 DB 28 FGYSKAFVSVLANIADHGHNVTF-QPPH-----IALKNVDGLINKNIE- 74  
 QY 69 IIAAY 73  
 DB 74 IINY 78

RESULT 11  
 ID 066004 PRELIMINARY; PRT: 102 AA.  
 AC 066004;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
 DT 01-AUG-1998 (TREMblrel. 07, Last annotation update)  
 DE MANNITOL-1-PHOSPHATE DEHYDROGENASE.  
 GN MTLN.  
 OS Mycoplasma bovine group 7.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molllicutes;  
 OC Mycoplasmataceae; Mycoplasma.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PG50;  
 RA MEDLINE: 98439347.  
 RA FREY J., CHENG X., MONNERAT M.-P., ABDO E.-M., KRAMINKLER M.,  
 RA BOLSE G., NICOLET J.;  
 RT "Genetic and serological analysis of the immunogenic 67-kDa  
 RT lipoprotein of Mycoplasma sp. bovine group 7.";  
 RL Res. Microbiol. 149:55-64(1998).

DR EMBL: U67071; AAC06132.1;  
 SQ SEQUENCE 102 AA; 12119 MW; 28F1AC19 CRC32;

Query Match 15.1%; Score 61.5; DB 2; Length 102;  
 Best Local Similarity 24.3%; Pred. No. 3.2;  
 Matches 18; Conservative 15; Mismatches 36; Indels 5; Gaps 1;

QY 2 KVTAFNPGPNRRVEIKRVSNI-HGRRIDIFASKNFHLQNTIGTGRISLTKTSKI 61  
 DB 25 KYNLSNLENDYKIQIKRLNSSEFKDDKRLVNTLKLKSN-----EKLITLDYKV 79  
 QY 62 ASRRVDGIIAAYQ 75  
 DB 80 SNLKHDTLLSLYQN 93

RESULT 12  
 ID 047446 PRELIMINARY; PRT: 173 AA.  
 AC 047446;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
 DE PAPE PROTEIN.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 88169520.  
 RA LUND B., LINDBERG F., NORMARK S.;  
 RT "Structure and antigenic properties of the tip-located P pilus  
 RT proteins of uropathogenic Escherichia coli.";  
 RL J. Bacteriol. 170:1887-1894(1988).  
 DR EMBL: M20181; AAA24291.1;  
 DR PFAM: PF00419; Fimbrial; 1.  
 DR SEQUENCE 173 AA; 18450 MW; E09D027D CRC32;

Query Match 15.0%; Score 61; DB 2; Length 173;  
 Best Local Similarity 27.0%; Pred. No. 7;  
 Matches 20; Conservative 11; Mismatches 35; Indels 8; Gaps 3;  
 QY 1 MKVTY-AFNGPGRNRRVEIKRVSNI-HGRRIDIFASKNFHLQNTIGTGRISLTKTS 59  
 DB 79 MKVTYATNTY--NNALIVNTSNTSSDGLVLYLNS-----MAGNIGTATLTGPTTPG 131  
 QY 60 KIASRRVDGIIAAY 73  
 DB 132 KITGNMADRTISLH 145

RESULT 13  
 ID 015152 PRELIMINARY; PRT: 726 AA.  
 AC 015152; 000645;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
 DE PLAKOPHILIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SCHMIDT A., HANS H.W., SCHAEFER S., NOBER U.A., ZIMBELMANN R.,  
 RA FRANK W.W.;  
 RL Submitted (Jul-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA ZIMBELMANN R.;  
 RL Submitted (Jul-1994) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.  
 RA BOSCH A.;  
 RI Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 SEQUENCE FROM N.A.  
 RA SCHMIDT A., LANGBEIN L., RODE M., PRAETZEL S., FRANK W.W.;  
 RI J. Mol. Biol. 0:0-0(0).  
 DR EMBL: 234974; CAAB4426.1; -  
 DR EMBL: 273678; CAAB8022.1; -  
 DR PFAM: PF00514; Armadillo\_seg. 3.  
 SEQUENCE 726 AA; 80496 MW; A92E30E9 CRC32;

Query Match 14.8%; Score 60; DB 4; Length 726;  
 Best Local Similarity 30.8%; Pred. No. 52;  
 Matches 12; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

QY 31 IDIFASKNFHLQKNTIGTGRISIKLTSGLASRRVDGI 69  
 DB 293 VDLRSPNQNVQAAAGALRNLFVSTNKLSTRONGI 331

RESULT 14  
 ID 013835 PRELIMINARY; PRT; 725 AA.  
 AC 013835;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE BAND-6-PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-EPIDERMIS;  
 RX MEDLINE: 95074299.  
 RA HATZFELD M., KRISTJANSSON G.I., PLESSMANN U., WEBER K.;  
 RT "Band 6 protein, a major constituent of desmosomes from stratified  
 epithelia, is a novel member of the armadillo multigene family."  
 RL J. Cell Sci. 107:2259-2270(1994).  
 DR EMBL: X79293; CA455881.1; -  
 DR PFAM: PF00514; Armadillo\_seg. 3.  
 SO SEQUENCE 725 AA; 80433 MW; 270BE242 CRC32;

Query Match 14.8%; Score 60; DB 4; Length 725;  
 Best Local Similarity 30.8%; Pred. No. 52;  
 Matches 12; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

QY 31 IDIFASKNFHLQKNTIGTGRISIKLTSGLASRRVDGI 69  
 DB 292 VDLRSPNQNVQAAAGALRNLFVSTNKLSTRONGI 330

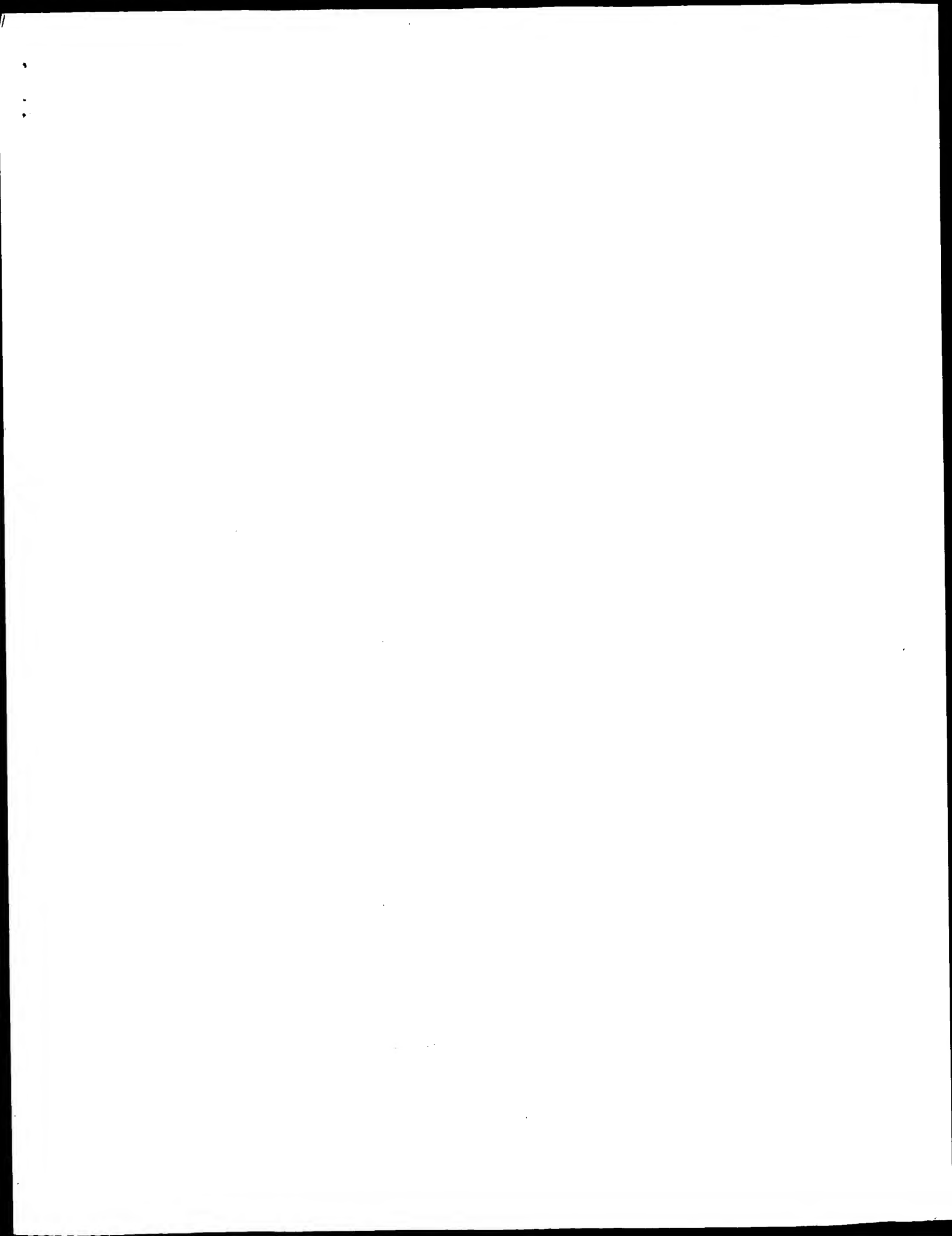
RESULT 15  
 ID 044584 PRELIMINARY; PRT; 275 AA.  
 AC 044584;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE F4867.4 PROTEIN.  
 GN F4867.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 CC Rhabditina; Rhabditoides; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans."  
 RL Nature 366:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA CLARKE K., WOHLDMANN P., HARRISON M.;  
 RI Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF039044; AF047944.1; -  
 SO SEQUENCE 275 AA; 32128 MW; E9B82465 CRC32;

Query Match 14.8%; Score 60; DB 5; Length 775;  
 Best Local Similarity 34.3%; Pred. No. 16;  
 Matches 12; Conservative 10; Mismatches 9; Indels 4; Gaps 1;

QY 36 SKNFHLQKNTIGTGRISIKLTSGLASRRVDGI 70  
 DB 230 SKRFYOKLQISTGERVLIV---KLSADTIEGVV 260

Search completed: September 29, 1999, 13:36:45  
 Job time: 2085 sec



OM of: US-09-142-524-1 to: GenEmbl.\* out-format : pfs

Date: Sep 29, 1999 1:59 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

#### Command line parameters:

-MODEL=frame+pn.model -DEV=rlp  
-O=/gen2/\_us09142524/runat\_29091999\_125958\_12782/arp\_query.fasta.1  
-UB=GenEmbl -OFMT=fastap -SUFFIX=rg -GAP=12.000 -GAPEXT=4.500  
-MATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000 -GAPOP=6.000  
-MATCH=0.050 -XGAPCL=10.000 -XGAPEXT=0.500 -EGAPOP=6.000  
-MATCH=7.000 -YGAPCL=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-MATCH=7.000 -STAR=1 -MATRIX=blomum62 -TRANS=human40.cdi  
-157-15 -DOCALLIGN=200 -THR\_SCORE=esscore -ALIGN=15 -MODE=LOCAL  
-OFMT=pfs -NOR=stat -USER=US09142524 -NCPU=6 -ICPU=3 -WAIT

#### Search Information block:

Query: US-09-142-524-1  
Query length: 80  
Database: GenEmbl.\*  
Database sequences: 679419  
Database length: 1590154680  
Search time (sec): 583.430000

#### score\_list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation
em.pat:E09607	+	99.00	202.96	0.0023	1545
em.pat:E09664	+	99.00	202.96	0.0023	1545
gb.pl1:CPOCRV11	+	98.00	202.96	0.0026	1708
em.pat:E10716	+	98.00	201.90	0.0026	1733
gb.pl1:CPOCRV1P	+	99.00	201.61	0.0027	1790
em.pat:E11677	+	74.00	149.69	2.13	1062
gb.pl1:D45404	+	74.00	148.16	2.28	1125
em.pat:E11679	+	74.00	148.13	2.61	1260
gb.pl1:DM85758	+	72.00	147.85	2.93	712
gb.to:RRRS18A	+	70.00	147.24	2.93	592
gb.to:WUSKE3B	+	70.00	147.08	2.99	525
gb.pl2:AF106662	+	73.00	147.07	2.99	1104
gb.pl2:AF106663	+	73.00	147.07	2.99	1104
gb.om:AB000911	+	70.00	147.03	3.01	528
gb.pr2:HSRPS18	+	70.00	146.67	3.15	549
em.pat:E11681	+	74.00	145.00	3.31	1542
em.pat:E11681	+	74.00	145.00	3.31	1542
gb.pl1:CPOCRV1B	+	71.00	141.35	6.22	1256
gb.pl1:CPOCRV1A	+	71.00	141.35	6.23	1257
gb.pl1:CPOCRV1I	+	71.00	141.29	6.27	1264
gb.pat:E07875	+	71.00	140.92	6.58	1317
gb.pr2:HS1N110P2	+	60.50	137.33	10.43	145
gb.sts:DM12C12S	+	62.50	134.44	15.10	326
gb.ov:AF051365	+	63.50	134.44	24.50	630
gb.pr2:HS1GMAAC	+	61.00	127.63	36.19	473
gb.ov:STJ97027	+	59.50	126.46	42.21	372
gb.pr2:HS243334	+	66.00	126.30	42.88	1883
gb.sts:G28131	+	60.00	125.96	44.79	567
gb.in1:DRORPS18A	+	61.00	125.96	44.79	567
gb.om:SSU15448	+	59.00	123.92	58.20	432
gb.pr2:HS1GMAAG	+	61.00	123.56	60.96	732
gb.pr2:HS1GMAAH	+	61.00	123.56	60.96	732
gb.ov:AF034264	+	63.50	122.96	65.80	737
gb.sts:Y1L24339	+	59.50	122.96	68.14	1400
gb.ov:Y1L24339	+	64.50	122.28	71.84	2015
gb.pr2:HS1GJV4	+	57.50	121.82	76.24	375
gb.pat:MC167	+	57.50	121.41	80.32	392
gb.pat:AS1518	+	65.00	121.05	84.12	2608
gb.pat:195550	+	65.00	121.05	84.12	2608
gb.pl1:YSCSECI7P	+	62.50	121.04	84.19	1406

gb.pat:AE7730 + 50.00 121.01 84.53 64 ! A67730 Sequence 60 from Paten  
gb.sts:G48233 + 58.00 120.74 87.47 477 ! G48233 SHGC-60275 Human Homo  
gb.pat:ECOPAPEF + 62.00 120.74 87.52 1284 ! M13239 E.Coli pape and pape

seq\_name: em.pat:E09607

seq\_documentation\_block:  
ID E09607 standard: RNA: UNC; 1545 BP.

AC E09607:

SV E09607.1

NI d1107944

DT 08-OCT-1997 (Rel. 52, Created)

DT 08-OCT-1997 (Rel. 52, Last updated, Version 1)

DE CDNA encoding polypeptide causing pollen allergy.

XX JP 1995170986-A/1.

XX unclassified

XX unclassified.

XX [1]

RA Namba M., Torigoe K., Kurimoto M.;

RT "DNA CODING POLYPEPTIDE, RECOMBINANT DNA CONTAINING THE DNA AND

RL Patent number JP 1995170986-A/1, 11-JUL-1995.

RL HAYASHIBARA BIOCHEM LAB INC.

OS Cryptomeria (cedar)

PN JP 1995170986-A/1

CC PD 11-JUL-1995

CC PF 20-DEC-1993 JP 1993344596

CC PR 05-NOV-1993 JP 93P 299151

CC PI NAKAMA MOTOJI, TORIGOE KAKUJI, KURIMOTO MASASHI

CC PC C12N15/09, C12N1/21, G01N33/50, A61K39/36, G01N33/53, G01N33/53,

CC PC C12N1/21;

CC CC strandness: Double;

CC CC topology: linear;

CC CC hypothetical: No;

CC CC anti-sense: No;

CC FH key Location/Qualifiers

CC FH source 1.1545

CC FH /db\_xref="taxon:32644"

CC FH /organism="Cryptomeria"

CC FH /tissue-type="pollen"

CC FH /product="polypeptide causing pollen allergy"

CC FH key Location/Qualifiers

CC FH source 1.1545

CC FH /db\_xref="taxon:32644"

CC FH /organism="unidentified"

SO Sequence 1545 BP; 508 A; 283 C; 352 G; 402 T; 0 other;

alignment\_scores:

Quality: 99.00

Ratio: 5.211

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-09-142-524-1 x E09607

Align seg 1/1 to: E09607 from: 1 to: 1545

31 ILEASPILEPHEALASERLYSAsnPHEHISLENGLINLYSAsnThrIleGI 47  
 |||||||  
 706 ATTGATATCTTTCATCTATAAAACTTTCATCTTCACAAAGAACACGATAG 755

47 ythrngly 49  
 |||||||  
 756 AACAGGG 762

seq\_name: em\_pat:E09664

seq\_documentation\_block:  
 ID E09664 standard; RNA; UNC; 1545 BP.

AC E09664;  
 XX  
 SV E09664.1  
 XX  
 NT d1108001

07-OCT-1997 (Rel. 52, Created)  
 07-OCT-1997 (Rel. 52, Last updated, Version 1)

CUWA encoding a novel protein which induces hay fever.

JP 1995188289-A/1.

unclassified.

[1]  
 RP 1-1545  
 RA Torigoe K., Namba M., Kurimoto M.;  
 RT "POLYPEPTIDE, ITS PRODUCTION AND USE";  
 RL Patent number JP 1995188289-A/1, 25-JUL-1995.  
 RL HAYASHIBARA BIOCHEM LAB INC.

XX OS Cryptomeria sp.  
 CC PN JP 1995188289-A/1  
 CC PD 25-JUL-1995  
 CC PF 27-DEC-1993 JP 1993346814  
 CC PI TORIGOE KAKUJI, NAMBA MOTOJI, KURIMOTO MASASHI  
 CC PC C07K14/415,A61K39/36,C12N15/09,C12P21/02,(C12P21/02,C12R1.19);  
 CC CC strandedness: Double;  
 CC CC topology: Linear;  
 CC FH Key Location/Qualifiers  
 CC FT source 1..1545  
 CC FT /organism="Cryptomeria sp."  
 CC FT /tissue\_type="pollen"  
 CC FT CDS 1..1545  
 CC FT /product="hay fever-inducing protein"  
 CC XX  
 FH Key Location/Qualifiers  
 FH FT source 1..1545  
 FH FT /db\_xref="taxon:32644"  
 FH FT /organism="unidentified"  
 FH FT  
 SO Sequence 1545 BP; 508 A; 281 C; 354 G; 402 T; 0 other;

alignment\_scores:

Quality: 99.00 Length: 19  
 Ratio: 5.211 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-142-524-1 x E09664 ..  
 Align seg 1/1 to: E09664 from: 1 to: 1545

31 ILEASPILEPHEALASERLYSAsnPHEHISLENGLINLYSAsnThrIleGI 47  
 |||||||

706 ATTGATATCTTTCATCTATAAAACTTTCATCTTCACAAAGAACACGATAG 755

47 ythrngly 49  
 |||||||  
 756 AACAGGG 762

seq\_name: gb\_p11:CPOCRYJ11

seq\_documentation\_block:

LOCUS CPOCRYJ11 1708 bp mRNA PLN 07-FEB-1999  
 DEFINITION Japanese cedar mRNA for Cry j II allergen, complete cds.  
 ACCESSION D29772  
 D29772  
 NID 9506857  
 VERSION D29772.1 GI:506857  
 KEYWORDS Cry j II; allergen.  
 SOURCE Cryptomeria japonica cDNA to mRNA.  
 ORGANISM Cryptomeria japonica  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;  
 Taxodiaceae; Cryptomeria.  
 1 (bases 1 to 1708)  
 Komiyama, N.

REFERENCE  
 AUTHORS  
 JOURNAL  
 TITLE

Submitted (02-APR-1994) to the DDBJ/EMBL/GenBank databases, Naoki Komiyama, Meiji Institute of Health Science, 540 Naruda, Odawara, Kanagawa 250, Japan (Tel:0465-37-3661, Fax:0465-36-2776)

REFERENCE  
 AUTHORS  
 JOURNAL  
 TITLE

Submitted (02-APR-1994) to the DDBJ/EMBL/GenBank databases, Naoki Komiyama, Meiji Institute of Health Science, 540 Naruda, Odawara, Kanagawa 250, Japan (Tel:0465-37-3661, Fax:0465-36-2776)

COMMENT  
 Submitted (02-Apr-1994) to DDBJ by:  
 Naoki Komiyama  
 Meiji Institute of Health Science  
 540 Naruda, Odawara  
 Kanagawa 250  
 Japan  
 Phone: 0465-37-3661  
 Fax: 0465-36-2776.  
 Location/Qualifiers  
 1..1708  
 /organism="Cryptomeria japonica"  
 /db\_xref="taxon:3369"  
 45..1589  
 /standard\_name="Cry j II"  
 /function="second major allergen of Japanese cedar pollen"  
 /codon\_start=1  
 /product="allergen"  
 /protein\_id="BA06172.1"  
 /db\_xref="PID:0106734"  
 /db\_xref="PID:9506858"  
 /db\_xref="GI:506858"  
 /translation="MAMKLIAPMAFLAQLIIMAAEDSOAIMLDSVEKYSNRS  
 LKVEHSRDAITNIFVEKYGAVGDKDCDEASTAWCAKNDPAALVPGSKFY  
 VNKLFRNGCOPHFTFKYDGIATAIQNPASKNKNTIWFQKLGFTLMGSGYIDGG  
 KQWMAQCKWVNGREICNDROPRVIAIKFDESTGLIQGLKANSPEHLYVGCNCGYK  
 IIGISTITAPRDSPNTDIDIFASKNFHLQKNTIGIDCVAILGSSNIVLEDCGP  
 GGISISIGRENSRAEVSIVVNGAKFIDIQNGLRITWGGSGMASHIIEYENEMI  
 NSENPILINOFCTGSASACQONRSVVOIODYTKNIKGTSAIAAIQKCDSPCKD  
 IRLSDSLIKTSKGTASCIQNDNANGFEGHVPACKNLSPAKRKRESKSHKPTVMY  
 ENMRATDKNRRIRILLGSRPPTCKKCGSCCKAKLYIVIRIMQEYYPGRWICSCSH  
 GRVHP"

FEATURES

source

CDS

BASE COUNT 568 a 294 c 380 g 466 t  
 ORIGIN

alignment\_scores:

Quality: 99.00 Length: 19  
 Ratio: 5.211 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-142-524-1 x CPOCRYJII ..

Align seg 1/1 to: CPOCRYJII from: 1 to: 1708

31 IleaapllephalaSerlysaSnPhenHisLeuGlnlysaSnThrIleGI 47  
 |||||||  
 750 ATGTGATATCTTGCATCTTAATAAACTTCACTTACAAAGAACACGATAGG 799

47 YThrGly 49  
 |||||||  
 800 AACAGCG 806

seq\_name: em\_pat:E10716

seq\_documentation\_block:

ID E10716 standard: RNA: UNC: 1733 BP.

AC E10716:

SV E10716.1

XX d1109053

XX 08-OCT-1997 (Rel. 52, Created)

XX 08-OCT-1997 (Rel. 52, last updated, Version 1)

XX cDNA encoding cedar pollen allergen, Cryj2.

XX JP 1996047392-A/1.

XX unidentified

XX unclassified.

XX [1]

XX 1-1733

XX Sone T., Komiyama N., Kii K.;

XX "ALLERGEN CRY J II EPI TOPE OF POLLEN OF JAPANESE CEDAR";

XX Patent number JP 1996047392-A/1. 20-FEB-1996.

XX MEIJI MILK PROD CO LTD.

XX OS Cryptomeria sp. (Cedar)

XX JP 1996047392-A/1

XX PD 20-FEB-1996

XX PR 05-NOV-1993 JP 93P 276773, 26-MAY-1994 JP 94P 134868

XX PI SOME TOSHIO, KOMIYAMA NAOKI, KII KOUSUKE

XX PC C12N15/09.C07K7/08.C07K14/415//A61K39/36.C12Q1/68.G01N33/53;

XX CC strandedness: Double;

XX CC topology: Linear;

XX CC hypothetical: No;

XX CC anti-sense: No;

XX CC Key

XX CC Location/Qualifiers

XX CC source

XX CC 1. 1733

XX CC /db\_xref="taxon:32644"

XX CC /organism="unidentified"

XX CC Sequence 1733 BP: 593 A; 294 C; 380 G; 466 T; 0 other;

alignment\_scores:

Quality: 99.00

Ratio: 5.211

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-09-142-524-1 x E10716 ..

Align seg 1/1 to: E10716 from: 1 to: 1733

31 IleaapllephalaSerlysaSnPhenHisLeuGlnlysaSnThrIleGI 47

|||||

750 ATGTGATATCTTGCATCTTAATAAACTTCACTTACAAAGAACACGATAGG 799

47 YThrGly 49

|||||

800 AACAGCG 806

seq\_name: gb\_p11:CPOCJIP

seq\_documentation\_block:

LOCUS CPOCJIP 1790 bp mRNA

DEFINITION Cryptomeria japonica mRNA for Cry j II, complete cds.

ACCESSION D37765

NID 9377695

VERSION D37765.1 GI:577695

KEYWORDS Cry j II; allergen.

SOURCE Cryptomeria japonica

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;

Taxodiaceae; Cryptomeria.

1 (bases 1 to 1790)

Namba, M.

Direct Submission

Submitted (22-JUL-1994) to the DDBJ/EMBL/GenBank databases. Motoshi

Namba, Fujisaki Institute, Hayashibara Biochemical

Laboratories, Inc; Fujisaki 675-1, Okayama, Okayama 702, Japan

(Tel:086-276-3141, Fax:086-276-6885)

2 (bases 1 to 1790)

Namba, M., Kurose, M., Torigoe, K., Hino, K., Taniguchi, Y., Fukuda, S.,

Usui, M. and Kurimoto, M.

Molecular cloning of the second major allergen, Cry j II, from

Japanese cedar pollen

FEBS Lett. 353 (2), 124-128 (1994)

95010777

FEATURES

MEDLINE

JOURNAL

TITLE

AUTHORS

REFERENCE

CDS

1. 1790

/organism="Cryptomeria japonica"

/db\_xref="taxon:3369"

/tissue\_type="pollen"

126..1670

/product="Cry j II precursor"

/protein\_id="BA407021.1"

/db\_xref="PID:d1007598"

/db\_xref="PID:g577696"

/translation="MAMKFLAPVAFVAMOLIIIMAAEDSOAIMDSDIEQVIRENRS

LRKVEHSRDAIINIEVVEKYGAVGCKHCTAFTAWDAAKKRSAMLVGNKKFV

VNNLFENGPCQPFRTKVDGIIAAYONPASMKNRRIWIOFALITPTLMGKVIDGOG

KOWMAQOCQWVNGREICNDNRDPTAIKFDSTGLIIQGLKLNSEFHLVFGNCEGVK

IIGISTAPRDSNTDGDIDIFASKNFKLQNTIGTGDVVAIGTSGSSNIIYEDLIGCP

GHGISTGSLGRENRAEVSIVYVNGAKFIDTONGARIKTWOGSGMSAIIYENEMI

NSENPIILINOFYCTISACONRSAYQIDVYKNIRGNRSATRAAIIOLKSGSMCKD

IKLSDISILKTSGLKASCINDNANGVFSCHVYPAKKNISPSAKRRESHSKRPKYVYV

KMGAYDKGNRRIRILIGSRPPTCKNCHGCSCKAKLVIVHINPOEYIPORWCSSRH

GKIYHP"

BASE COUNT 585 a 325 c 392 g 488 t

ORIGIN

## alignment\_scores:

Quality: 99.00 Length: 19  
Ratio: 5.211 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-142-524-1 x CPOCJIP ..

Align seg 1/1 to: CPOCJIP from: 1 to: 1790

31 IIEAEPHlephelaSerLysAsnPhenHisLeuGlnLysAsnThrIleG1 47  
|||||  
831 ATGATATCTTGCATCTTAACCTTCACTTACAAAGAACACGATAG 880  
47 ythrGly 49  
|||||  
881 AACAGCG 887

seq\_name: em\_pat:E11677

seq\_documentation\_block:  
ID E11677 Standard; RNA; UNC; 1062 BP.

XX AC E11677;  
XX SV E11677.1  
XX NI d1110014  
XX DT 08-OCT-1997 (Rel. 52, Created)  
XX DT 08-OCT-1997 (Rel. 52, Last updated, Version 1)  
XX DE cDNA encoding Cha o I.  
XX KW JP 1996176192-A/1.  
XX OS unidentified  
XX OC unclassified.  
XX RN [1]  
XX RP 1-1062  
XX RA Sone T., Komiyama N., Kii K.;  
XX RT "ALLERGEN TO POLLEN OF CHAMAECYPARIS OBTUSA";  
XX RL Patent number JP 1996176192-A/1, 09-JUL-1996.  
XX RL MEITI MILK PROD CO LTD.

XX OS Chamaecyparis obtusa  
XX CC PN JP 1996176192-A/1  
XX CC PD 09-JUL-1996  
XX CC PF 21-DEC-1994 JP 1994335089  
XX PI SONE TOSHIO, KOMIYAMA NAOKI, KII KOUSUKE  
XX PC C07K14/415,C12N15/09,C12P21/02//A61K35/12,A61K35/64,A61K35/72,  
XX PC A61K35/74,A61K39/36;  
XX CC PC A61K35/78,A61K39/36;  
XX CC CC strandedness: Double;  
XX CC CC topology: Linear;  
XX CC FH key Location/Qualifiers  
XX CC FT source 1..1062  
XX CC FT /organism="Chamaecyparis obtusa"  
XX CC FT /tissue\_type="pollen"  
XX CC FT mat\_peptide 1..1062  
XX CC FT /product="Cha o I"

XX FH key Location/Qualifiers

XX FH source

XX FH 1..1062  
XX FH /db\_xref="taxon:32644"  
XX FH /organism="unidentified"

XX SQ Sequence 1062 BP; 319 A; 198 C; 254 G; 291 T; 0 other;

## alignment\_scores:

Quality: 74.00 Length: 27  
Ratio: 3.895 Gaps: 0  
Percent Similarity: 70.370 Percent Identity: 59.259

alignment\_block:  
US-09-142-524-1 x E11677 ..

Align seg 1/1 to: E11677 from: 1 to: 1062

1 MetLysValThrValAlaPheAsnGlnPheGlyProAsnArgValPh 17  
|||||  
637 ATGACGTGACAGTGGCATTCATCAATTTGGACCTAATGCTGACAACG 686  
17 eileLysArgValSerAsnValIleIleHis 27  
::: ||::: ::|||  
687 AATGCCAGGCGCAGTATGACTATATACAT 717

seq\_name: em\_pat:E11678

seq\_documentation\_block:  
ID E11678 Standard; RNA; UNC; 1125 BP.

XX AC E11678;  
XX SV E11678.1  
XX NI d1110015  
XX DT 08-OCT-1997 (Rel. 52, Created)  
XX DT 08-OCT-1997 (Rel. 52, Last updated, Version 1)  
XX DE cDNA encoding Cha o I precursor.  
XX KW JP 1996176192-A/2.  
XX OS unidentified  
XX OC unclassified.  
XX RN [1]  
XX RP 1-1125  
XX RA Sone T., Komiyama N., Kii K.;  
XX RT "ALLERGEN TO POLLEN OF CHAMAECYPARIS OBTUSA";  
XX RL Patent number JP 1996176192-A/2, 09-JUL-1996.  
XX RL MEITI MILK PROD CO LTD.

XX OS Chamaecyparis obtusa  
XX CC PN JP 1996176192-A/2  
XX CC PD 09-JUL-1996  
XX CC PF 21-DEC-1994 JP 1994335089  
XX PI SONE TOSHIO, KOMIYAMA NAOKI, KII KOUSUKE  
XX PC C07K14/415,C12N15/09,C12P21/02//A61K35/12,A61K35/64,A61K35/72,  
XX PC A61K35/74,A61K39/36;  
XX CC PC A61K35/78,A61K39/36;  
XX CC CC strandedness: Double;  
XX CC CC topology: Linear;  
XX CC FH key Location/Qualifiers  
XX CC FT source 1..1125  
XX CC FT /organism="Chamaecyparis obtusa"  
XX CC FT /tissue\_type="pollen"  
XX CC FT sig\_peptide 1..63  
XX CC FT mat\_peptide 64..1125  
XX CC FT /product="Cha o I"

XX FH key Location/Qualifiers

XX FH source

XX FH 1..1125  
XX FH /db\_xref="taxon:32644"  
XX FH /organism="unidentified"

XX SQ Sequence 1125 BP; 327 A; 211 C; 264 G; 323 T; 0 other;

VMLGHSIDTVSDDKSRKVTAFNFGNGNGRMPBRARVYLIIHANNNDPWSYIAIGG  
 GSVPRAGKODDATHRNATVDWVMDHNSDSDGLVLTASTGYLISNNHFHNRHNR  
 MPLYIAGKTDITRGAEVHIGNGGCLFMRTVSHYILBLNTHICNTSVSGVNLISEL  
 /VLFSSAEEM--WASCLLVAFLFCLVASCSDNPIDSCWGDAMWDMNMRKLD  
 CAGGSSAMGKGAGGGLVATVSDDDPVPAGTLCRYGMRRESMILFEKNIKIKL  
 PLVAGNATDITRGAEVHIGNGGCLFMRTVSHYILBLNTHICNTSVSGVNLISEL

SC FT sig

Accession	Source	Accession	Source
CC	1..1260	/organism="Chamaecyparis obtusa"	
CC		/tissue_type="pollen"	
CC		50..1177	
CC		/product="Cha o I precursor"	
CC		50..112	
FT	sig_peptide		

changes in mated female flies"

/codon.start=1  
/product="accessory gland protein Acp29AB"  
/protein\_id="AAB96382.1"  
/db\_xref="PID:g2772591"  
/db\_xref="GI:2772591"  
/translation="MYASMLYLALIMNLMDLSGGODIPNGKATLPSPTPONTIDQ  
IGIONVYFNALKNQETLAIIDEMEMIASLLEFPAQHEIDLOPKIMRHASN  
IKASNNIMKRFKVSRRHHEIKENIMOTFAAYTCRMNGSHLANIDOEKELDLILA  
LAPNNSWIDISIKLVNNGGFFVSTLGRPFPPVKKNSODIKKKNOGVYIYAKEMSYD  
EEFESKSFVCAQADMA"

BASE COUNT 282 a 155 c 162 g 193 t

ORIGIN

alignment\_scores:  
Quality: 72.00 Length: 81  
Ratio: 1.532 Gaps: 3  
Percent Similarity: 58.025 Percent Identity: 25.926

alignment\_block:  
US-09-142-524-1 x DM085758 ..

Align seg 1/1 to: DM085758 from: 1 to: 792

9 GlnPhgelyprobsnargarGVALPheilelysarGVALSerAsnVal.. 24  
||||: ||| ::::: ::::: |||||:::|  
290 CAGCTTCAGCCGTTAAAGATTATAAAGCACACCATGCTCCACATCAA 339  
25 .....IlelleHISGLYArgArglleAspIlePhealasertLysa 38  
||| ||| |||||:::|  
340 AGCCTTACACACATCAAGATGAGAGCATTCGAGAAAGTTGGCTCCAGAC 389  
38 snPheHISLeuGlnLysAsnThrIleGlyThr.....GlyArgArgIle 52  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
390 ATTTCACATCGAGAGAAATCTATACCAACTGGTTGAGGCATATATGTC 439  
53 SerLeuLysLeuThrSerGlyLysIleAlaser.....Argar 65  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
440 ACATGTCGTAATAATGAAACGGTCATCTGCGCAACATCCAGAGTGAAGA 489  
65 gVALAspGlylleIleAlaAlaTyrglnAsnProAlaserTTP 79  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
490 GCTGATGCGCATCTTGGCGTTAGCACCCACACATCAATCACTG 532

seq\_name: gb\_ro:RRPS18a

seq\_documentation\_block:  
LOCUS RRP518A 516 bp mRNA ROD 30-NOV-1993  
DEFINITION R.rattus ribosomal protein S18 mRNA.  
ACCESSION X57529 S51078  
NID 9433446  
VERSION X57529.1 GI:433446  
KEYWORDS ribosomal protein S18.  
SOURCE black rat.  
ORGANISM Rattus rattus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 516)  
Chan, Y.L.  
Direct Submission  
Submitted (06-FEB-1991) Y.L. Chan, University of Chicago, 920 East  
58th Street, Chicago IL 60637, U S A  
2 (bases 1 to 516)  
Chan, Y.L., Paz, V. and Wool, I.G.  
The primary structure of rat ribosomal protein S18  
Biochem. Biophys. Res. Commun. 178 (3), 1212-1218 (1991)  
91337062 Location/Qualifiers  
1..516  
/organism="Rattus rattus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10117"

FEATURES  
SOURCE

CDS

38..496  
/clone="pmc313-1"  
/codon\_start=1  
/product="ribosomal protein S18"  
/protein\_id="CAA40750.1"  
/db\_xref="GI:433447"  
/db\_xref="GI:433447"  
/db\_xref="SWISS-PROT:P25232"  
/translation="MSLVPEKFOHILRVLTNTIDGRRIKIAFAITAKGVRRYAHV  
LKRADIDLRRAEELTEDEVERVITIMQNPQYKIPDWFILNRQKVDGKYSQVLANG  
LDNKLREDELRKIRAHRGIRHFWGLRVGQHTKTGTGRGRITGVSKK"

BASE COUNT 127 a 133 c 159 g 97 t

ORIGIN

alignment\_scores:  
Quality: 70.00 Length: 79  
Ratio: 1.707 Gaps: 2  
Percent Similarity: 51.899 Percent Identity: 30.380

alignment\_block:  
US-09-142-524-1 x RRP518A ..

Align seg 1/1 to: RRP518A from: 1 to: 516

```

12 ProSnaArgArgValPheIleLysArgValSerAsnValIleIleHISGI 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 CCCGAGAGTTTCAGACATCTCGAGTCTCAGACCAACATCGATG 102
28 YARGArgIleAspIlePheAlaSerLysAsnPheHisLeuGlnLysAsnT 45
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 GCGGCGGAAATAGCGCTTGCCTATCAGTCCATTAAG..... 139
45 hrIleGlyThrGlyArgArg..... 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
140 .....GGTGTGGGGCGGAGATACCGTCATGTGTGAGGAAAGCAGAC 184
52 IleSerLeuLysLeuThrSerGlyLysIleAlaSerArgArgValAspGI 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 ATTGACCTCACCAAGAGGCTGGGAGAGCTCAGCAGACGAGAGTGAGCG 234
68 YIleIleAlaAlaIleTyrGlnAsnProAlaSerTrpLys 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 TGTGATCACCATCATGACAGACCCAGACAAATACAAAG 271

```

seq\_name: gb\_ro:MUSKE3B

seq\_documentation\_block:  
LOCUS MUSKE3B 525 bp mRNA ROD 16-MAR-1994  
DEFINITION Mus musculus ribosomal protein (Ke-3) mRNA, complete cds.  
ACCESSION M76763  
M76763.1 GI:198579  
VERSION 9198579  
KEYWORDS ribosomal protein; ribosomal protein S13 homologue.  
SOURCE Mus musculus cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 525)  
AUTHORS MacMurray,A.J. and Shin,H.S.  
TITLE The murine MHC encodes a mammalian homolog of bacterial ribosomal protein S13  
JOURNAL Mamm. Genome 2 (2), 87-95 (1992)  
MEDLINE 92182530  
FEATURES  
Source Location/Qualifiers  
1..525  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
1..22  
/partial  
/note="putative"  
23..525  
/gene="Ke-3"

CDS

23..481  
/gene="Ke-3"  
/note="putative"  
/codon\_start=1  
/product="ribosomal protein"  
/protein\_id="AA16795.1"  
/db\_xref="GI:198580"  
/db\_xref="GI:198580"  
/translation="MSLVPEKFOHILRVLTNTIDGRRIKIAFAITAKGVRRYAHV  
LKRADIDLRRAEELTEDEVERVITIMQNPQYKIPDWFILNRQKVDGKYSQVLANG  
LDNKLREDELRKIRAHRGIRHFWGLRVGQHTKTGTGRGRITGVSKK"

3 UTR

polyA\_signal  
502..507  
/gene="Ke-3"  
/note="putative"

polyA\_site  
525  
/gene="Ke-3"  
/note="putative"

BASE COUNT 138 a 128 c 156 g 103 t

ORIGIN

alignment\_scores:  
Quality: 70.00 Length: 79  
Ratio: 1.707 Gaps: 2  
Percent Similarity: 51.899 Percent Identity: 30.380

alignment\_block:  
US-09-142-524-1 x MUSKE3B ..

Align seg 1/1 to: MUSKE3B from: 1 to: 525

```

12 ProSnaArgArgValPheIleLysArgValSerAsnValIleIleHISGI 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 CTGAGAGAGTTCCAGACATTTTGGAGTACTCAGACCAACATCGATG 87
28 YARGArgIleAspIlePheAlaSerLysAsnPheHisLeuGlnLysAsnT 45
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 GCGGCGGAAATAGCGCTTGCCTATCAGTCCATTAAG..... 124
45 hrIleGlyThrGlyArgArg..... 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 .....GCGTGGGGCGGAGATACCGTCATGTGTGAGGAAAGCAGAC 169
52 IleSerLeuLysLeuThrSerGlyLysIleAlaSerArgArgValAspGI 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 ATGACCTCACCAAGAGGCTGGGAGAGCTCAGGAGGATGAGTGAGCG 219
68 YIleIleAlaAlaIleTyrGlnAsnProAlaSerTrpLys 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 AGTGATCACCATCATGACAGACCCAGACGACACTACAG 256

```

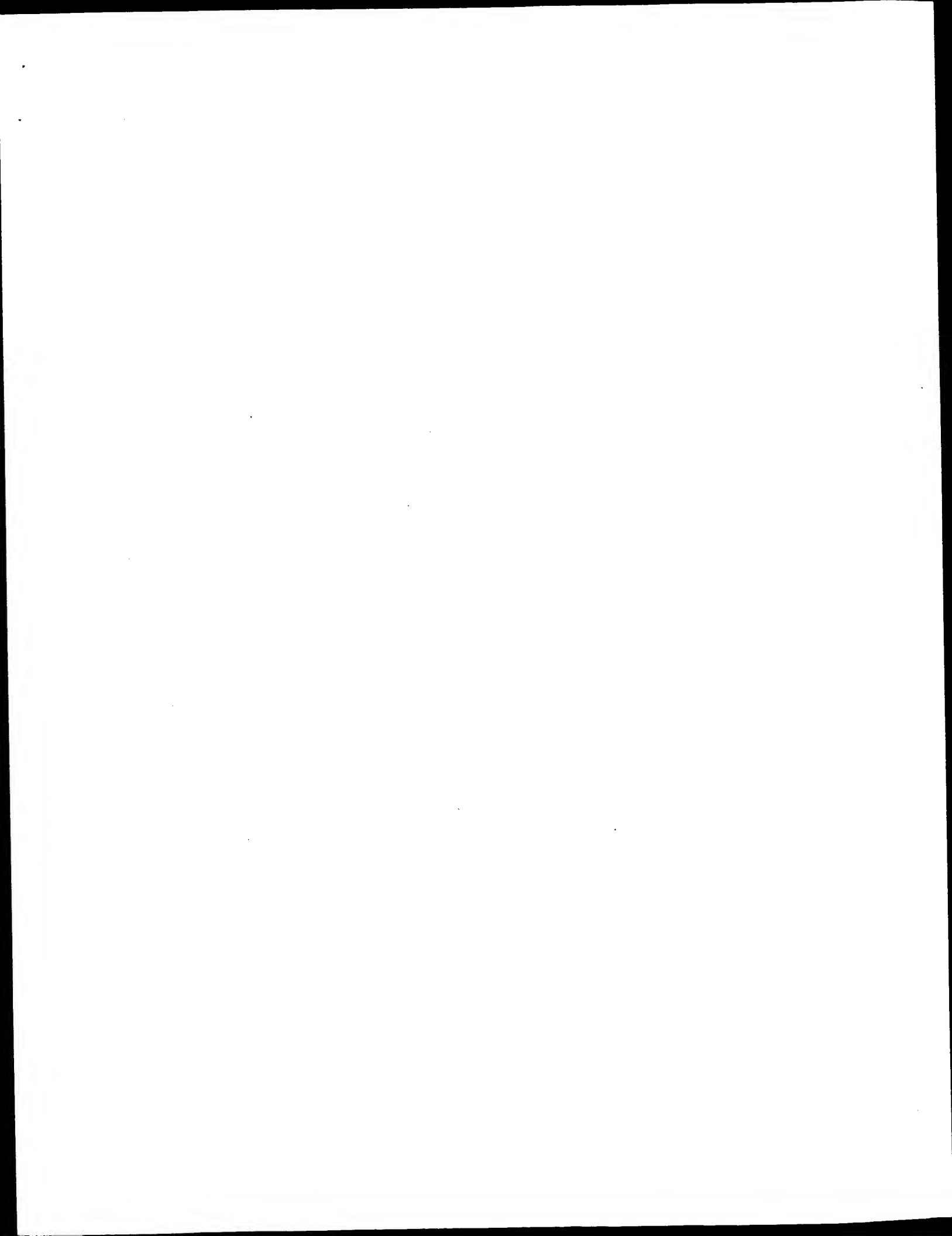
seq\_name: gb\_PL2:AF106662

seq\_documentation\_block:  
LOCUS AF106662 1104 bp mRNA PLN 10-JAN-1999  
DEFINITION Juniperus ashei pollen major allergen 1-1 mRNA, complete cds.  
ACCESSION AF106662  
AF106662.1 GI:4138876  
VERSION 94138876  
KEYWORDS allergen; Juniperus.  
SOURCE Ozark white cedar.  
ORGANISM Juniperus ashei  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;  
Cupressaceae; Juniperus.  
REFERENCE 1 (bases 1 to 1104)  
AUTHORS Midoro-Horiuchi,T.M., Goldblum,R.M., Kurosky,A., Wood,T.G. and Brooks,E.G.  
TITLE Molecular cloning of mountain cedar (Juniperus ashei) pollen major allergen, Jun a 1

JOURNAL Unpublished  
 2 (bases 1 to 1104)  
 Mido-ro-Horiuti, T.M., Goldblum, R.M., Kurosky, A., Wood, T.G. and  
 Brooks, E.G.  
 TITLE Direct Submission  
 Submitted (12-NOV-1998) Department of Pediatrics, Child Health  
 Research Center, University of Texas Medical Branch, 301 University  
 Blvd., Galveston, TX 77555-0366, USA  
 FEATURES Location/Qualifiers  
 source 1..1104  
 /organism="Juniperus ashei"  
 /db\_xref="taxon:13101"  
 1..1104  
 /note="Jun a 1-1"  
 CDS  
 /codon\_start=1  
 /product="pollen major allergen 1-1"  
 /protein\_id="A003608.1"  
 /db\_xref="PID:g4138879"  
 /db\_xref="GI:4138877"  
 /translation="MASPCLIAVIVFCAIVSCYSDNPIDSCMRGDSNMNDONRMLAD  
 CAVGSSSTMGGKGGDEFTVSTDDNPNVPTPTGTRIGATREKALMIFISQNMNLIK  
 MLYVAGHKTIDGRGADVHLGNGPCLEFMKRVSHVLIHSGNTSVLGDLVSES  
 IGVEPVHAGDQDAITMRNVTNAMDHNSLSDCSDGLIDVTLSGTITISNHFENHKK  
 VMLGHDPTDYDDKSMKVTVAENFGPNAGORMPRARVGLVHANNNDPNNIYAIG  
 SNNPTILSEGNSTAPSESYKKEVTKRIGCESPSACANWVWRSTDAFINGAVFVSSG  
 KTEETNITNSNEAFKVENGNAPOLTKNAGVVT"  
 BASE COUNT 324 a 202 c 264 g 314 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 73.00 Length: 27  
 Ratio: 3.842 Gaps: 0  
 Percent Similarity: 70.370 Percent Identity: 55.556  
 alignment\_block:  
 US-09-142-524-1 x AF106662 ..  
 Align seg 1/1 to: AF106662 from: 1 to: 1104  
 1 MettysValThrValAlaIlePheAsnGlnPheGlyProAsnArgValAlaP 17  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 700 ATGAAAGTGACAGTGGCGTCAATTCATGACCTAATGCTGGCGCAAG 749  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 17 eilelYsArgValSerAsnValIlelleHs 27  
 ::|||::|  
 750 AATGCCAAGGCGACGATGACTTGATCAT 780  
 seq\_name: gb\_p12:AF106663  
 seq\_documentation\_block:  
 LOCUS AF106663 1104 bp mRNA PLN 10-JAN-1999  
 DEFINITION Juniperus ashei pollen major allergen 1-2 mRNA, complete cds.  
 ACCESSION AF106663  
 NID 94138878  
 VERSION AF106663.1 GI:4138878  
 KEYWORDS  
 ORGANISM  
 Juniperus ashei  
 Ozark white cedar.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;  
 Cupressaceae; Juniperus.  
 REFERENCE  
 1 (bases 1 to 1104)  
 Mido-ro-Horiuti, T.M., Goldblum, R.M., Kurosky, A., Wood, T.G. and  
 Brooks, E.G.  
 TITLE Molecular cloning of mountain cedar (Juniperus ashei) pollen major  
 allergen, Jun a 1  
 JOURNAL Unpublished  
 2 (bases 1 to 1104)  
 Mido-ro-Horiuti, T.M., Goldblum, R.M., Kurosky, A., Wood, T.G. and  
 Brooks, E.G.  
 TITLE Direct Submission  
 Submitted (12-NOV-1998) Department of Pediatrics, Child Health

Research Center, University of Texas Medical Branch, 301 University  
 Blvd., Galveston, TX 77555-0366, USA  
 FEATURES Location/Qualifiers  
 source 1..1104  
 /organism="Juniperus ashei"  
 /db\_xref="taxon:13101"  
 1..1104  
 /note="Jun a 1-2"  
 CDS  
 /codon\_start=1  
 /product="pollen major allergen 1-2"  
 /protein\_id="A003609.1"  
 /db\_xref="PID:g4138879"  
 /db\_xref="GI:4138879"  
 /translation="MASPCLIAVIVFCAIVSCYSDNPIDSCMRGDSNMNDONRMLAD  
 CAVGSSSTMGGKGGDEFTVSTDDNPNVPTPTGTRIGATREKALMIFISQNMNLIK  
 MLYVAGHKTIDGRGADVHLGNGPCLEFMKRVSHVLIHSGNTSVLGDLVSES  
 IGVEPVHAGDQDAITMRNVTNAMDHNSLSDCSDGLIDVTLSGTITISNHFENHKK  
 VMLGHDPTDYDDKSMKVTVAENFGPNAGORMPRARVGLVHANNNDPNNIYAIG  
 SNNPTILSEGNSTAPSESYKKEVTKRIGCESPSACANWVWRSTDAFINGAVFVSSG  
 KTEETNITNSNEAFKVENGNAPOLTKNAGVVT"  
 BASE COUNT 324 a 203 c 264 g 313 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 73.00 Length: 27  
 Ratio: 3.842 Gaps: 0  
 Percent Similarity: 70.370 Percent Identity: 55.556  
 alignment\_block:  
 US-09-142-524-1 x AF106663 ..  
 Align seg 1/1 to: AF106663 from: 1 to: 1104  
 1 MettysValThrValAlaIlePheAsnGlnPheGlyProAsnArgValAlaP 17  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 700 ATGAAAGTGACAGTGGCGTCAATTCATGACCTAATGCTGGCGCAAG 749  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 17 eilelYsArgValSerAsnValIlelleHs 27  
 ::|||::|  
 750 AATGCCAAGGCGACGATGACTTGATCAT 780  
 seq\_name: gb\_cm:AB000911  
 seq\_documentation\_block:  
 LOCUS AB000911 528 bp mRNA MAM 05-FEB-1999  
 DEFINITION Sus scrofa mRNA for ribosomal protein, complete cds.  
 ACCESSION AB000911  
 NID 91841303  
 VERSION AB000911.1 GI:1841303  
 KEYWORDS ribosomal protein S13.  
 SOURCE Sus scrofa CDNA to mRNA.  
 ORGANISM  
 Sus scrofa  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Artiodactyla; Suidae; Sus.  
 REFERENCE  
 1 (bases 1 to 528)  
 Hamamatsu, N.  
 TITLE Direct Submission  
 Submitted (07-FEB-1997) to the DDBJ/EMBL/GenBank databases.  
 Noriyuki Hamamatsu, STAF-Institute, Animal Genome Research Program  
 Team; 446-1 Ippaiyuka, Kamiyokoda, Tsukuba, Ibaraki 305, Japan  
 (E-mail:hamamatsu@genome.staif.or.jp, Tel:0298-38-2190,  
 Fax:0298-38-2337)  
 2 (sites)  
 Kimura, M., Kawakami, K., Suzuki, H. and Hamamatsu, N.  
 TITLE Cloning of the pig homolog of bacterial ribosomal protein S13  
 JOURNAL Unpublished (1997)  
 FEATURES Location/Qualifiers  
 source 1..528  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 30..488  
 /codon\_start=1







PT Allergenic Cry j II protein and fragments from Japanese cedar  
 TT pollen - used to diagnose, treat and prevent Japanese cedar  
 PT pollinosis  
 PS Disclosure: Page 49; 89pp; English.  
 CC The sequence is of a Japanese cedar pollen allergen Cry j IIA  
 CC fragment. The DNA and its fragments can be used for diagnosis and  
 CC treatment of Japanese cedar pollinosis and to identify similar  
 CC sequences in other plants.  
 CC See also Q66048-67.  
 SQ Sequence 1395 BP; 460 A; 257 C; 318 G; 360 T;

alignment\_scores:  
 Quality: 99.00 Length: 19  
 Ratio: 5.211 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-142-524-1 x Q66050 ..

Align seg 1/1 to: Q66050 from: 1 to: 1395

31 ILEASPILEPHEALASERLYSASNPHEHISLEUGLNLYSASNTHTLLEGI 47  
 |||||||  
 556 ATTGATATCTTTCGATCTAAACCTTCACTTACAAAGAACACGATAGG 605  
 47 ythr gly 49  
 |||||||  
 606 AACAGGG 612

seq\_name: N\_Geneseq\_36:Q66049

seq\_documentation\_block:

ID Q66049 standard; DNA: 1410 BP.  
 AC Q66049;  
 DT 01-FEB-1995 (first entry)  
 DE Japanese cedar pollen allergen Cry j II DNA fragment.  
 KM Cedar pollinosis; diagnostic; ss.  
 OS Crytomeria japonica.  
 PN WO9411512-A.  
 PD 26-MAY-1994.  
 PE 12-NOV-1993; U11000.  
 PR 12-NOV-1992; US-975179.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Brauer A, Kuo M, Pollock J, Yeung S;  
 DR WPI: 94-183513/22.  
 PT Allergenic Cry j II protein and fragments from Japanese cedar  
 PT pollen - used to diagnose, treat and prevent Japanese cedar  
 PT pollinosis  
 PS Disclosure: Page 48; 89pp; English.  
 CC The sequence is of a Japanese cedar pollen allergen Cry j IIA  
 CC fragment. The DNA and its fragments can be used for diagnosis and  
 CC treatment of Japanese cedar pollinosis and to identify similar  
 CC sequences in other plants.  
 CC See also Q66048-67.  
 SQ Sequence 1410 BP; 467 A; 258 C; 322 G; 363 T;

alignment\_scores:  
 Quality: 99.00 Length: 19  
 Ratio: 5.211 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-142-524-1 x Q66049 ..

Align seg 1/1 to: Q66049 from: 1 to: 1410

31 ILEASPILEPHEALASERLYSASNPHEHISLEUGLNLYSASNTHTLLEGI 47  
 |||||||  
 571 ATTGATATCTTTCGATCTAAACCTTCACTTACAAAGAACACGATAGG 620  
 47 ythr gly 49

||||||  
 621 AACAGGG 627

seq\_name: N\_Geneseq\_36:Q66051

seq\_documentation\_block:

ID Q66051 standard; DNA: 1479 BP.  
 AC Q66051;  
 DT 01-FEB-1995 (first entry)  
 DE Japanese cedar pollen allergen Cry j II DNA fragment.  
 KM Cedar pollinosis; diagnostic; ss.  
 OS Crytomeria japonica.  
 PN WO9411512-A.  
 PD 26-MAY-1994.  
 PE 12-NOV-1993; U11000.  
 PR 12-NOV-1992; US-975179.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Brauer A, Kuo M, Pollock J, Yeung S;  
 DR WPI: 94-183513/22.  
 PT Allergenic Cry j II protein and fragments from Japanese cedar  
 PT pollen - used to diagnose, treat and prevent Japanese cedar  
 PT pollinosis  
 PS Disclosure: Page 50; 89pp; English.  
 CC The sequence is of a Japanese cedar pollen allergen Cry j IIA  
 CC fragment. The DNA and its fragments can be used for diagnosis and  
 CC treatment of Japanese cedar pollinosis and to identify similar  
 CC sequences in other plants.  
 CC See also Q66048-67.  
 SQ Sequence 1479 BP; 490 A; 268 C; 337 G; 384 T;

alignment\_scores:  
 Quality: 99.00 Length: 19  
 Ratio: 5.211 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-142-524-1 x Q66051 ..

Align seg 1/1 to: Q66051 from: 1 to: 1479

31 ILEASPILEPHEALASERLYSASNPHEHISLEUGLNLYSASNTHTLLEGI 47  
 |||||||  
 640 ATTGATATCTTTCGATCTAAACCTTCACTTACAAAGAACACGATAGG 689  
 47 ythr gly 49  
 |||||||  
 690 AACAGGG 696

seq\_name: N\_Geneseq\_36:Q84045

seq\_documentation\_block:

ID Q84045 standard; cDNA: 1542 BP.  
 AC Q84045;  
 DT 27-SEP-1995 (first entry)  
 DE Japonicum allergen cDNA.  
 KM Japonicum allergen; induced histamine release; antiallergic peptide;  
 KW IgE cross-linking inhibition; ds.  
 OS Japonicum sp.  
 FH Key Location/Qualifiers  
 FT mat\_peptide 1..1542  
 FT /\*tag= a  
 PN WO9502412-A.  
 PD 26-JAN-1995.  
 PE 15-JUL-1994; J01164.  
 PR 16-JUL-1993; JP-177008.  
 PR 01-SEP-1993; JP-217725.  
 PR 07-APR-1994; JP-069336.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Kino K, Kohno Y, Komiyama N, Sone T;  
 DR WPI: 95-067159/09.  
 DR P-PDB: R69792.  
 PT Peptide antiallergic agent - inhibits cross-linking of allergen

PT With IGE antibody  
 PS Disclosure: Pages 27-28: 46pp: Japanese.  
 CC Q84045 encodes R69792 Japonicum allergen, from which the antiallergic  
 CC peptides R69845-R69809 were derived. The peptides ability to inhibit  
 CC the cross-linking of an allergen, to an IGE antibody can be used in  
 SO the prevention and treatment of allergic diseases.  
 Sequence 1542 BP: 506 A: 280 C: 352 G: 404 T;

alignment\_scores:  
 Quality: 99.00 Length: 19  
 Ratio: 5.211 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-142-524-1 x Q84045 ..

Align seg 1/1 to: Q84045 from: 1 to: 1542

31 IleaapllephelaSerlysaSnphenHisleuGlnlysaSnhrlllegl 47  
 |||||||  
 706 ATTGATATCTTTCATCTTAATAACTTTCACCTTACAAAGAACAGATAGG 755  
 |||||||  
 47 yThrcly 49  
 |||||||  
 756 AACACGG 762

seq\_name: N\_Geneseq\_36:Q90156

seq\_documentation\_block:

ID Q90156 standard; DNA: 1545 BP.

AC Q90156;

DT 01-NOV-1995 (first entry)

DE Japanese cedar pollen allergen gene.

KW Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;

OS Cryptomeria japonica

FH Key Location/Qualifiers

FT cds 1..1545

FT /\*tag= a

PN EP-655500-A.

PD 31-MAY-1995.

PE 03-NOV-1994: 308117.

PR 05-NOV-1993: JP-299151.

PR 20-DEC-1993: JP-344596.

PR 27-DEC-1993: JP-346814.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PI Kurimoto M, Namba M, Torisgoe K;

DR P-PSDB: R74333.

PT New Japanese cedar pollen allergen polypeptide - and DNA coding for it,

PS useful for treatment and diagnosis of cedar pollen allergy

CC Claim 6; Page 29; 41pp; English.

CC The gene encoding an allergen of Japanese cedar pollen was isolated

CC by PCR amplification using primers based on portions of the allergen

CC protein. The gene was used for recombinant allergen production in

CC E. coli (vector plasmid pCR-223-3).

CC Sequence 1545 BP: 508 A: 283 C: 351 G: 403 T;

seq\_name: N\_Geneseq\_36:Q90156

seq\_documentation\_block:

ID Q84045 standard; CDNA: 1733 BP.

AC Q84045;

DT 27-SEP-1995 (first entry)

DE Japonicum allergen CDNA and flanking sequences.

KW Japonicum allergen; induced histamine release; antiallergic peptide;

OS Japonicum sp.

FH Key Location/Qualifiers

FT cds 45..1589

FT /\*tag= a

PN MO9502412-A.

PD 26-JAN-1995.

PE 15-JUL-1994: J01164.

PR 16-JUL-1993: JP-177008.

PR 01-SEP-1993: JP-217725.

47 yThrcly 49  
 |||||||  
 756 AACACGG 762

seq\_name: N\_Geneseq\_36:Q66048

seq\_documentation\_block:

ID Q66048 standard; CDNA: 1726 BP.

AC Q66048;

DT 01-FEB-1995 (first entry)

DE Japanese cedar pollen allergen Cry j II.

KW Cedar pollinosis; diagnostic; ss.

OS Cryptomeria japonica.

FH Key Location/Qualifiers

FT cds 42..1586

FT /\*tag= a

PN MO9411512-A.

PD 26-MAY-1994.

PE 12-NOV-1993: U11000.

PR 12-NOV-1992: US-975179.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Brauer A, Kuo M, Pollock J, Yeung S;

DR WPI: 94-183513/22.

DR P-PSDB: R53690.

PT Allergenic Cry j II protein and fragments from Japanese cedar

PT pollen - used to diagnose, treat and prevent Japanese cedar

PT pollinosis

PS Claim 2; Fig 4: 89pp; English.

CC The sequence is that encoding a Japanese cedar pollen allergen Cry j

CC II. The DNA and its fragments can be used for diagnosis and

CC treatment of Japanese cedar pollinosis and to identify similar

CC sequences in other plants.

CC See also Q66049-67.

SO Sequence 1726 BP: 579 A: 299 C: 383 G: 465 T;

alignment\_scores:

Quality: 99.00 Length: 19

Ratio: 5.211 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-142-524-1 x Q66048 ..

Align seg 1/1 to: Q66048 from: 1 to: 1726

31 IleaapllephelaSerlysaSnphenHisleuGlnlysaSnhrlllegl 47  
 |||||||  
 747 ATTGATATCTTTCATCTTAATAACTTTCACCTTACAAAGAACAGATAGG 796  
 |||||||  
 47 yThrcly 49  
 |||||||  
 797 AACACGG 803

seq\_name: N\_Geneseq\_36:Q84046

seq\_documentation\_block:

ID Q84046 standard; CDNA: 1733 BP.

AC Q84046;

DT 27-SEP-1995 (first entry)

DE Japonicum allergen CDNA and flanking sequences.

KW Japonicum allergen; induced histamine release; antiallergic peptide;

OS Japonicum sp.

FH Key Location/Qualifiers

FT cds 45..1589

FT /\*tag= a

PN MO9502412-A.

PD 26-JAN-1995.

PE 15-JUL-1994: J01164.

PR 16-JUL-1993: JP-177008.

PR 01-SEP-1993: JP-217725.

PR 07-APR-1994; JP-0693336.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI KINO K, Kohno Y, Komiyama N, Sone T;  
 DR WPI: 95-067159/09.  
 DR P-PSDB; R69792.  
 PT Peptide antiallergic agent - inhibits cross-linking of allergen  
 with IgE antibody  
 PS Example 1: Page 29; 46pp; Japanese.  
 CC 084046 encodes R69792 Japonicum allergen, from which the antiallergic  
 CC peptides R69845-R69809 were derived. The peptides ability to inhibit  
 CC the cross-linking of an allergen, to an IgE antibody can be used in  
 CC the prevention and treatment of allergic diseases.  
 SQ Sequence 1733 BP; 593 A; 294 C; 380 G; 466 T;

alignment\_scores:  
 Quality: 99.00 Length: 19  
 Ratio: 5.211 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-142-524-1 x 084046 ..

Align seg 1/1 to: 084046 from: 1 to: 1733

31 ILeAspIlePheAlaSerLysAsnPhenHisLeuGlnLysAsnThrIleGI 47  
 |||||  
 750 ATGTGATATCTTGGCATCTAAACCTTCACCTTACAAAGAACAGCATAGG 799  
 47 yThrGly 49  
 |||||  
 800 AACACGGG 806

seq\_name: N\_Geneseq\_36:T18102

seq\_documentation\_block:

ID T18102 standard; cDNA to mRNA; 1733 BP.

AC T18102:

DI 16-AUG-1996 (first entry)

DE Japan cedar pollen cDNA encoding Cry j II allergen.

KW Sugil pollen; epitope; overlapping peptide; Cry j II; cedar pollen;

OS Cryptomeria japonica.

PT Key Location/Qualifiers

FT 5'utr 1..44

FT cds /tag= a

FT signal\_peptide 45..206

FT mat\_peptide 207..1586

FT 3'utr 1590..1733

FT /tag= d

FT /tag= e

PN J08047392-A.

PD 20-FEB-1996.

PF 07-NOV-1994; 297840.

PI US-NOV-1993; JP-276773.

PR 26-MAY-1994; JP-134868.

PA (MEIP) MEIJI MILK PROD CO LTD.

PI WPI: 96-166249/17.

DR P-PSDB; R93599.

PT Japan cedar pollen allergen Cry j II epitope - comprises at least

part of specified 460 amino acid protein

PS Claim 2: Page 13-14; 17pp; Japanese.

CC T18102 encodes a Japan cedar pollen Cry j II allergen which

is useful in the diagnosis, prevention and treatment of sugil

pollinosis, the allergic reaction to Japan cedar pollen. Significant

regions of the allergen were identified using overlapping peptides

of the full epitope derived from a Cry j II antigen specific T cell

line (see R97871-R97960). Amino acids 66-80 (R97884) and 186-200

CC (R978908) of the full mature 460 amino acid allergen are the most

allergenic of the 90 peptides tested.

SQ Sequence 1733 BP; 593 A; 294 C; 380 G; 466 T;  
 alignment\_scores:  
 Quality: 99.00 Length: 19  
 Ratio: 5.211 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-142-524-1 x T18102 ..

Align seg 1/1 to: T18102 from: 1 to: 1733

31 ILeAspIlePheAlaSerLysAsnPhenHisLeuGlnLysAsnThrIleGI 47  
 |||||  
 750 ATGTGATATCTTGGCATCTAAACCTTCACCTTACAAAGAACAGCATAGG 799  
 47 yThrGly 49  
 |||||  
 800 AACACGGG 806

seq\_name: N\_Geneseq\_36:T38518

seq\_documentation\_block:

ID T38518 standard; cDNA; 1125 BP.

AC T38518:

DI 28-NOV-1996 (first entry)

DE Chamaecyparis obtusa pollen allergen Cha o I cDNA (B).

KW Pollen allergen; Cha o I; T-cell epitope; prevention; treatment;

OS Chamaecyparis obtusa.

PT Key Location/Qualifiers

FT cds 1..1125

FT /tag= a

FT /note= "STOP codon absent"

PN J08176192-A.

PD 09-JUL-1996.

PF 21-DEC-1994; 335089.

PI 21-DEC-1994; JP-335089.

PA (MEIP) MEIJI MILK PROD CO LTD.

DR WPI: 96-368225/37.

DR P-PSDB; W04345.

PT DNA encoding chamaecyparis obtusa pollen allergen - T cell

epitope(s) of which are useful in development of preventative and

treating agent for C. obtusa pollen pollinosis

PS Claim 10; Page 13; 17pp; Japanese.

CC The present sequence encodes the C. obtusa pollen allergen Cha o I.

CC the T-cell epitopes of which can be used in the development of a

CC preventive and treating agent for C. obtusa pollen pollinosis.

CC C. obtusa pollen (2.4 kg) was degreased with diethyl ether, and

CC dried at room temp. overnight. Cha o I was sepd. from it and

CC purified. RNA was extracted from C. obtusa pollen, and mRNA and

CC cDNA derived.

CC Sequence 1125 BP; 327 A; 211 C; 264 G; 323 T;

alignment\_scores:

Quality: 74.00 Length: 27

Ratio: 3.895 Gaps: 0

Percent Similarity: 70.370 Percent Identity: 59.259

alignment\_block:

US-09-142-524-1 x T38518 ..

Align seg 1/1 to: T38518 from: 1 to: 1125

1 MetLysValThrValAlaPheAsnGlnPheGlyProAsnArgValPhe 17  
 |||||  
 700 ATGAAGGTCAGTCAGTCATCAATCAATTTGACCTAATGCTGCACACG 749  
 17 eLLeYsArgValSerAsnValIleIleHis 27  
 |||||  
 750 AATGCCAAGGCGACGATATGACCTATATCAT 780

```

seq_name: N_Geneseq_36:T38519
seq_documentation_block:
ID T38519 standard; cDNA; 1260 BP.
AC T38519;
DE 28-NOV-1996 (first entry)
DE Chamaecyparis obtusa pollen allergen Cha o I cDNA (A).
KW Pollen allergen; Cha o I; T-cell epitope; prevention; treatment;
OS Chamaecyparis obtusa.
FH Key Location/Qualifiers
FT cds 113..1177
FT /tag= a
FT /note= "sequence comprising nucleotides 113..1174
PN J08176192-A.
PD 09-JUL-1996.
PF 21-DEC-1994; 335089.
PR 21-DEC-1994; JP-335089.
RA (MEIP) MEIJI MILK PROD CO LTD.
DR WPI; 96-368225/37.
DR P-PSDB; W04344.
CC DNA encoding chamaecyparis obtusa pollen allergen - T cell
CC epitope(s) of which are useful in development of preventative and
CC treating agent for C. obtusa pollen pollinosis
CC Claim 4; Pages 13-14; 17pp; Japanese.
CC The present sequence encodes the C. obtusa pollen allergen Cha o I,
CC the T-cell epitopes of which can be used in the development of a
CC preventive and treating agent for C. obtusa pollen pollinosis.
CC C. obtusa pollen (2.4 kg) was degassed with diethyl ether, and
CC dried at room temp. overnight. Cha o I was sepd. from it and
CC purified. RNA was extracted from C. obtusa pollen, and mRNA and
CC cDNA derived.
SQ Sequence 1260 BP; 375 A; 233 C; 282 G; 370 T;

alignment_scores:
Quality: 74.00 Length: 27
Ratio: 3.895 Gaps: 0
Percent Similarity: 70.370 Percent Identity: 59.259

alignment_block:
US-09-142-524-1 x T38519 ..
Align seg 1/1 to: T38519 from: 1 to: 1260

1 MetLysValThrValAlaIapheAnGlnPheGlyProAsnArgValAlp 17
|||||
749 ATGAAGTACAGTGCATTCATCAATTGGACCTAAATGCTGGACAACG 798
17 eileLysArgValSerAsnValIleIleHis 27
::: |||||
799 AATGCCAAGCGACGATATGACCTATATCAT 829

seq_name: N_Geneseq_36:Q55272
seq_documentation_block:
ID Q55272 standard; cDNA; 1170 BP.
AC Q55272;
DE 13-JUL-1994 (first entry)
DE Jun s I gene.
DE Japanese cedar; pollen allergen; allergy; treatment; diagnosis;
KW T cell epitope; sensitivity; detection; ss.
OS Juniperus sabinooides.
FH Key Location/Qualifiers
FT cds 26..1129
FT /tag= a
FT signal_peptide 26..88
FT /tag= b
FT mat_peptide 89..1129
FT /tag= c
PN W09401560-A.

```

```

PD 20-JAN-1994.
PF 15-JAN-1993; U00139.
PR 10-JUL-1992; WC-U05661.
PR 01-SEP-1992; US-938990.
RA (IMMO-) IMMUNOLOGIC PHARM CORP.
PI Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J;
DR WPI; 94-035066/04.
DR P-PSDB; R45577.
CC Antigen derived from Japanese cedar pollen allergen Cry j I -
CC contain at least two T cell epitope(s), used to treat or diagnose
CC allergy
PS Disclosure; Fig 16; 137pp; English.
CC The sequence is that encoding Jun s I, a homologue of the Japanese
CC cedar pollen allergen Cry j I. Antigenic peptides derived from it
CC can be used for the treatment and diagnosis of allergies associated
CC with Japanese cedar pollen.
SQ Sequence 1170 BP; 350 A; 215 C; 268 G; 337 T;

alignment_scores:
Quality: 73.00 Length: 27
Ratio: 3.842 Gaps: 0
Percent Similarity: 70.370 Percent Identity: 55.556

alignment_block:
US-09-142-524-1 x Q55272 ..
Align seg 1/1 to: Q55272 from: 1 to: 1170

1 MetLysValThrValAlaIapheAnGlnPheGlyProAsnArgValAlp 17
|||||
725 ATGAAGTACAGTGCATTCATCAATTGGACCTAAATGCTGGACAACG 774
17 eileLysArgValSerAsnValIleIleHis 27
::: |||||
775 AATGCCAAGCGACGATATGACCTTGTACAT 805

seq_name: N_Geneseq_36:T38521
seq_documentation_block:
ID T38521 standard; cDNA; 1772 BP.
AC T38521;
DE 28-NOV-1996 (first entry)
DE Chamaecyparis obtusa pollen allergen Cha o II cDNA.
KW Pollen allergen; Cha o II; T-cell epitope; prevention; treatment;
OS Chamaecyparis obtusa.
FH Key Location/Qualifiers
FT cds 32..1576
FT /tag= a
FT /note= "sequence comprising nucleotides 32..1573
FT /note= "sequence comprising nucleotides 32..1573
PN J08176192-A.
PD 09-JUL-1996.
PF 21-DEC-1994; 335089.
PR 21-DEC-1994; JP-335089.
RA (MEIP) MEIJI MILK PROD CO LTD.
DR WPI; 96-368225/37.
DR P-PSDB; W04344.
CC DNA encoding chamaecyparis obtusa pollen allergen - T cell
CC epitope(s) of which are useful in development of preventative and
CC treating agent for C. obtusa pollen pollinosis
CC Claim 16; Page 16; 17pp; Japanese.
CC The present sequence encodes the C. obtusa pollen allergen
CC Cha o II, the T-cell epitopes of which can be used in the
CC development of a preventative and treating agent for C. obtusa
CC pollen pollinosis; C. obtusa pollen (2.4 kg) was degassed with
CC diethyl ether, and dried at room temp. overnight. Cha o II was
CC sepd. from it and purified. RNA was extracted from C. obtusa
CC pollen, and mRNA and cDNA derived.
SQ Sequence 1772 BP; 581 A; 328 C; 388 G; 475 T;

```

alignment\_scores:                      length:                      19  
                     quality:                      74.00  
                     ratio:                      4.111  
 Percent Similarity:    94.737                      Percent Identity:    73.684

alignment\_block:  
 US-09-142-524-1 x T38521 ..

Align seg 1/1 to: T38521 from: 1 to: 1772

```

31 11leaplephelaaserlyasnphehisleuglnlyasnhrilegl 47
   |||||.....|
737 ATTGATATCTTTCATCTTAAGAATTCACATAGAAAGTCGTAATAGG 786
47 yThrGly 49
   |||||
787 AACACGG 793

```

seq\_name: N\_Geneseq\_36:071601

seq\_documentation\_block:  
 ID Q71601 standard; cDNA to mRNA; 1317 BP.

AC Q71601;  
 NT 24-MAR-1995 (first entry)  
 Japanese cedar pollen antigen Cryj1 cDNA.  
 Japanese cedar pollen antigen; allergen; Cryj1; sugi; pollinosis;

US Cryptomeria japonica.  
 Location/Qualifiers

```

..      signal_peptide 62..124
FT      mat_peptide 125..1183
FT      3'utr 1184..1317
FT      /*tag= d

```

```

PN J06197768-A.
PD 19-JUL-1994.
PE 07-JAN-1993; JP-001116.
PR 07-JAN-1993; JP-001116.
PA (MEIJ) MEIJU SEIKA KAISHA.
DR WPI; 94-268680/33.
DR P-PSDB; R60166.
PT Sugi (Japanese cedar) pollen antigen Cryj1 - is useful for
PS diagnosis, treatment and prevention of sugi pollinosis
CC Claim 6: Page 5-7; 9pp: Japanese.
CC The coding sequence for the Japanese cedar ("sugi") pollen allergen
CC Cryj1 was isolated from a cDNA library prepared from polyA mRNA. All
CC or part of the Cryj1 protein can be used for diagnosis, treatment
CC and prevention of sugi pollinosis.
SQ Sequence 1317 BP; 410 A; 230 C; 284 G; 393 T;

```

alignment\_scores:                      length:                      27  
                     quality:                      71.00  
                     ratio:                      3.737  
 Percent Similarity:    70.370                      Percent Identity:    55.556

alignment\_block:  
 US-09-142-524-1 x Q71601 ..

Align seg 1/1 to: Q71601 from: 1 to: 1317

```

1 MetLysValThrValAlaIapheasnGlnPheGlyProAsnArgArgValPh 17
   |||||.....|
761 ATGAAGGTGACAGTGGCGCTTCAATCATTTGACCTAACGTGGACAAAG 810
17 eileLysArgValSerAsnValIleIleHis 27
   ::|||:::
811 AATGCCCGACGACGATATGACTGTACAT 841

```

seq\_name: N\_Geneseq\_36:035304

seq\_documentation\_block:

ID Q35304 standard; cDNA to mRNA; 1337 BP.

AC Q35304;  
 NT 03-JUN-1993 (first entry)  
 DE Cry j 1 gene.  
 KW Japanese cedar pollen; allergen; antigen; allergy; B cell; T cell; ss.  
 OS Cryptomeria japonica.  
 FH Key location/Qualifiers  
 FT cds 66..1187  
 FT signal\_peptide 66..128  
 FT /\*tag= a  
 FT /\*tag= b  
 FT mat\_peptide 129..1187  
 FT /\*tag= c

```

PN M09301213-A.
PD 21-JAN-1993.
PE 10-JUL-1992; U05661.
PR 12-JUL-1991; US-728134.
PR 15-JUL-1991; US-730452.
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
PI Bond JF, Griffith J, Pollock J;
DR WPI; 93-045434/05.
DR P-PSDB; R31937.

```

PT Nucleic acid sequence encoding Cryptomeria japonica allergen -  
 for the diagnosis treatment and prevention of allergic reactions  
 to Japanese cedar pollen  
 PS Claim 1: Page 40; 69pp; English.  
 CC Fresh pollen and staminate cone samples were collected from a single  
 CC Cryptomeria japonica (Japanese cedar) tree. RNA was prep'd. and used  
 CC to synthesise cDNA. The cDNA was subjected to successive rounds of  
 CC PCR to yield a 193 bp clone, JCT1.6, found to contain a partial  
 CC sequence of Cry j 1. A secondary PCR was performed and clones for  
 CC full length Cry j 1, pC190C91a and pC190C91d (identical clones)  
 CC were identified. Cry j 1 or an antigenic fragment of it may be used  
 CC for detecting, treating and preventing an allergic response to  
 CC Japanese cedar pollen allergen. It is capable of modifying both the  
 CC B and T cell response to Cry j 1 and T cell response to a Cry j 1  
 CC antigen. The gene may be used to identify related allergens in other  
 CC plant species.  
 CC See also Q35305-20.  
 CC Sequence 1337 BP; 408 A; 236 C; 292 G; 401 T;

alignment\_scores:                      length:                      27  
                     quality:                      71.00  
                     ratio:                      3.737  
 Percent Similarity:    70.370                      Percent Identity:    55.556

alignment\_block:  
 US-09-142-524-1 x Q35304 ..

Align seg 1/1 to: Q35304 from: 1 to: 1337

```

1 MetLysValThrValAlaIapheasnGlnPheGlyProAsnArgArgValPh 17
   |||||.....|
765 ATGAAGGTGACAGTGGCGCTTCAATCATTTGACCTAACGTGGACAAAG 814
17 eileLysArgValSerAsnValIleIleHis 27
   ::|||:::
815 AATGCCCGACGACGATATGACTGTACAT 845

```

OM of: US-09-142-524-1 to: EST:\* out\_format : pfs

Date: Sep 29, 1999 1:50 PM

About: Results were produced by the GenCore software, version 4.5.  
Copyright (c) 1993-1998 Compugen Ltd.

# Command line parameters:

-MODEL=frame+22n.model -DEV=xlp  
-O=/cgn2.1/USPO.spool/US09142524/runat\_29091999.125958.12773/app\_query.fasta.1  
-DB=EST -QFMT=fastap -SUFFIX=1st -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -CGAPOP=4.500  
-CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCGAPOP=6.000  
-GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=DIOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=escore -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=stat -USER=US09142524 -NCPU=6 -ICPU=3 -WAIT  
-THREADS=1

## Search information block:

Query: US-09-142-524-1  
Query length: 80  
Database: EST:\*  
Database sequences: 2546578  
Database length: 96266752  
Search time (sec): 759.040000

## Score list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_est17:AA559604	80.00	193.24	0.0300	519	AA659604 nt63e10.s1 NCI CGAP_P3
gb_est11:AA498109	77.00	185.43	0.0818	513	AA498109 v175n08.r1 Stratagene
gb_est11:AA198386	77.00	185.30	0.0831	520	AA198386 mw48e09.r1 Soares mos
gb_est11:AA248242	74.00	180.84	0.1474	360	AA248242 csq1171.seq.f Human fe
gb_est12:AA036117	74.00	180.55	0.1528	371	AA036117 AU036117 polyandrocarr
gb_est12:AA138625	75.00	180.03	0.1655	520	AA138625 nt15e09.s1 NCI CGAP_GG
gb_est12:AA036104	73.00	179.30	0.1794	320	AA036104 AU036104 Polyandrocarr
gb_est12:AA300566	72.00	178.43	0.2006	265	AA300566 EST11420 Testis tumor
gb_est17:D76822	72.00	177.02	0.2404	308	D76822 MUR8A01 mouse embryonal
gb_est17:AA603354	72.00	176.31	0.2633	583	AA603354 hp06e11.s1 NCI CGAP_P3
gb_est17:AA372511	72.00	176.01	0.2766	343	AA372511 EST84502 Colon adenoca
gb_est14:AA381359	71.00	175.87	0.2788	268	AA381359 EST94431 Activated T-c
gb_est13:AA288847	71.00	175.69	0.2852	263	AA288847 m752c03.r1 Life Tech T
gb_est13:AA379609	71.00	175.31	0.2996	490	AA379609 EST94431 Activated T-c
gb_est13:AA378031	73.00	175.05	0.3097	287	AA378031 EST90680 Synovial sarc
gb_est13:AA815411	71.00	174.89	0.3099	504	AA815411 a161e03.s1 Soares test
gb_est13:AA349283	71.00	174.89	0.3162	292	AA349283 EST94431 Activated T-c
gb_est14:AA381326	71.00	174.74	0.3175	293	AA381326 EST94431 Activated T-c
gb_est14:AA381559	71.00	174.73	0.3228	293	AA381559 EST94431 Activated T-c
gb_est14:AA386415	71.00	174.69	0.3241	298	AA386415 EST94431 Activated T-c
gb_est19:AA792506	73.00	174.48	0.3329	230	AA792506 v92b06.r1 Batstead mo
gb_est12:AA300816	73.00	174.48	0.3330	535	AA300816 EST13940 Testis tumor
gb_est12:AA815996	73.00	174.46	0.3332	305	AA815996 v175n08.r1 Batstead mo
gb_est10:AA816017	73.00	174.43	0.3338	536	AA816017 v175n08.r1 Batstead mo
gb_est15:AA511865	73.00	174.32	0.3398	543	AA511865 v917c08.r1 Soares mos
gb_est13:AA358932	71.00	174.32	0.3398	310	AA358932 EST65553 Jurkat T-cell
gb_est13:AA379634	71.00	174.20	0.3451	314	AA379634 EST92728 Skin tumor I
gb_est13:AA330864	70.00	173.97	0.3504	318	AA330864 EST92728 Skin tumor I
gb_est13:AA334566	71.00	173.94	0.3526	243	AA334566 EST92728 Skin tumor I
gb_est13:AA33827	71.00	173.91	0.3570	243	AA33827 EST92728 Skin tumor I
gb_est13:AA376323	70.00	173.78	0.3584	323	AA376323 EST92728 Skin tumor I
gb_est12:AA321646	70.00	173.77	0.3644	248	AA321646 EST92728 Skin tumor I
gb_est13:AA53470	70.00	173.55	0.3650	329	AA53470 EST92728 Skin tumor I
gb_est14:AA381322	71.00	173.54	0.3751	254	AA381322 EST92728 Skin tumor I
gb_est13:AA379368	71.00	173.49	0.3757	337	AA379368 EST92728 Skin tumor I
gb_est13:AA360159	70.00	173.48	0.3786	339	AA360159 EST92728 Skin tumor I
gb_est13:AA353178	70.00	173.44	0.3804	257	AA353178 EST92728 Skin tumor I

gb\_est13:AA377911 + 71.00 173.43 0.3811 341 ! AA377911 EST90625 Synovial s  
gb\_est13:AA379646 + 71.00 173.43 0.3811 341 ! AA379646 EST92511 Skin tumor  
gb\_est14:AA381374 + 71.00 173.35 0.3852 344 ! AA381374 EST94447 Activated

seq\_name: gb\_est17:AA659604

## seq\_documentation\_block:

LOCUS AA659604 519 bp mRNA EST 05-NOV-1997  
DEFINITION nt63e10.s1 NCI CGAP\_P3 Homo sapiens CDNA clone IMAGE:1203210  
similar to gb:X69150.40S RIBOSOMAL PROTEIN S18 (HUMAN);, mRNA  
sequence.  
ACCESSION AA659604  
NID g2595758  
KEYWORDS AA659604.1 GI:2595758  
EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 519)  
NCI CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Unpublished (1997)  
On May 9, 1995 this sequence version replaced gi:802232.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuangui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMW at:  
[www.bio.limn.gov/btrp/image/image.html](http://www.bio.limn.gov/btrp/image/image.html)

Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 377.

Location/Qualifiers  
1. 519  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1203210"  
/clone\_lib="NCI CGAP\_P3"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"

/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected cells  
histologically-determined to be fully malignant prostate  
cancer cells. Double-stranded cDNA was ligated to EcoRI  
adaptor-specific primer, and the resulting PCR product  
subcloned into pAMP10 by the UDG-cloning method (Life  
Technologies). Average insert size is 600 bp. NOTE: Not  
directionally cloned. This library was constructed by  
David Krizman."

BASE COUNT 145 a 114 c 143 g 117 t

ORIGIN

alignment\_scores: 80.00 Length: 74  
Ratio: 2.000 Gaps: 2  
Percent Similarity: 54.054 Percent Identity: 32.432

alignment\_block: ..  
US-09-142-524-1 x AA659604

```

/organism="Mus musculus"
/strain="inbred CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:91811"
/clone_1ib="f1rr4tagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kannamycin resistant)"
note="Organ: testis; Vector: plasmidscript SK-, Site_1

```

Contact: Marra M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGS Consortium (info@image.llnl.gov) for further information.  
MGI:404160

Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 477.  
Location/Qualifiers  
1..520

FEATURES  
source  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:558312"  
/clone\_lib="Squares mouse 3NME12 5"  
/sex="unknown"  
/tissue\_type="fetus"  
/dev\_stage="12.5dpc total fetus"  
/lab\_host="DH10B"  
/note="Organ: whole fetus; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 3', on total mouse RNA (provided by Minoru Ko, Wayne State Univ.); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 137 a 127 c 153 g 103 t

ORIGIN

alignment\_scores:  
Quality: 77.00 Length: 88  
Ratio: 1.638 Gaps: 2  
Percent Similarity: 53.409 Percent Identity: 30.682

alignment\_block:  
US-09-142-524-1 x AA198386 ..

Align seg 1/1 to: AA198386 from: 1 to: 520

```

3 ValThValAlaphaennglnpheglyProasnArgArgValPheIlely 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 GTGACCGCGCGCATGTCTAGTATCGATCGAGAAATTCACACACTTTT 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 sArgValSerAsnValIleIleHsglyArgArgIleAspIlePheAlas 36
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 GCGAGTACACACACACATCATGATGCGCGGAAATAGCCCTCCGCCA 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
36 eRlyshnphHisLeuglnlyAsnThrIleGlyThrGlyArg... 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 TCACTGCCATTAG.....GCGGTGGGCGGAGATAT 141
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 .....IleSerLeuylsLeuThrSerg1 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 GCTCATGTGTTTGAGAAACAGACATCGACCTCACCAAGAGGCTGG 191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 yLysIleAlaSerArgArgValAspGlyIleIleAlaIleAlaTyrglnAsn 76
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 AAAATCAACGAGGATGAGGTGAGCGAGTATCATCATGACGAAAC 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 roAlaSerTrpLys 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 CACGACAGTACAAAG 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: gb_est11:AA248242

```

seq\_documentation\_block:  
LOCUS AA248242 360 bp mRNA EST 11-MAR-1997  
DEFINITION cs9117.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens  
CDNA 5', mRNA sequence.  
ACCESSION AA248242  
MID 91878895  
VERSION AA248242.1 GI:1878895  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 360)  
Liew/C.C.  
CDNAs from human fetal heart (1997)  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1402306.

Contact: Liew CC  
Department of Laboratory Medicine and Pathobiology  
University of Toronto  
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5  
Tel: 4169788758  
Fax: 4169785650  
Email: liewcc@utcc.utoronto.ca  
PCR Primers  
FORWARD: 5' GCCAGCTGAATTAACCCCTACTAAAGG 3'  
BACKWARD: 5' CCGATTAATGTATACGACTCAGCTATGAGGCG 3'  
Seq primer: 5' GAATTAACCTCATCAAGG 3'.  
Location/Qualifiers  
1..360  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="907F10"  
/clone\_lib="Human fetal heart, Lambda ZAP Express"  
/lab\_host="E. coli XL1-Blue"  
/note="Vector: Lambda ZAP Express; Site\_1: EcoRI; Site\_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested Lambda ZAP Express."

BASE COUNT 102 a 84 c 100 g 74 t

ORIGIN

alignment\_scores:  
Quality: 74.00 Length: 79  
Ratio: 1.805 Gaps: 2  
Percent Similarity: 51.899 Percent Identity: 31.646

alignment\_block:  
US-09-142-524-1 x AA248242 ..

Align seg 1/1 to: AA248242 from: 1 to: 360

```

12 ProAsnArgArgValPheIleLysArgValSerAsnValIleIleHsgl 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 CCTGAAGAGTCCAGCATATTTTCCAGTACTACACACACATCGATG 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28 yArgArgIleAspIlePheAlaSerLysAsnphHisLeuglnlyAsnT 45
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 GCGCGGAAATAGCCTTTCATCATGCTCATAG..... 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 hrIleGlyThrGlyArg... 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 .....GCTGGGCGGAGATATGCTCATGTGTTGAGAAACAGAC 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 IleSerLeuylsLeuThrSerg1yLysIleAlaSerArgArgValAspG 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 ATGACCTCACCAAGAGGCGGAGAGTACCTGAGAGAGTGTGACCG 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 yIleIleAlaIleAlaTyrglnAsnProAlaSerTrpLys 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 TGTGATCCACATTATGCAAGATTCACGCGCATGACAG 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: gb_est29:A0036117

```

seq\_documentation\_block:  
LOCUS A0036117 371 bp mRNA EST 26-MAR-1999  
DEFINITION A0036117 Polyandrocampa misakiensis white spot budding stage  
Polyandrocampa misakiensis cDNA, mRNA sequence.  
ACCESSION A0036117

NID 94527078  
 VERSION AU036117.1 GI:4527078  
 KEYWORDS EST.  
 SOURCE Polyandrocarpa misakiensis.  
 ORGANISM Polyandrocarpa misakiensis.  
 Eukaryota; Metazoa; Chordata; Polyandrocarpa; Stolidobranchia; Styelidae; Ascidiacea;

REFERENCE 1 (bases 1 to 371)  
 Kawamura, K., Hayata, D., Fujiwara, S. and Yubisui, T.  
 Serine protease inhibitors expressed in the process of budding of tunicates as revealed by EST analysis  
 J. Biochem. (1998) In press  
 On Jun 5, 1998 this sequence version replaced gi:3188168.

JOURNAL COMMENT

CONTACT: Kawamura K  
 Faculty of Science  
 Kochi University  
 2-5-1, Akebono-cho, Kochi 780, Japan  
 Tel: +81-888-44-8313  
 Fax: +81-888-44-8313  
 Email: kazukerc.kochi-u.ac.jp.

FEATURES  
 source location/Qualifiers  
 1..371  
 /organism="Polyandrocarpa misakiensis"  
 /strain="white spot"  
 /db\_xref="taxon:7723"  
 /clone\_lib="Polyandrocarpa misakiensis white spot budding stage"  
 /dev\_stage="budding stage"  
 /dev\_stage="budding stage"

BASE COUNT 103 a 83 c 100 g 84 t 1 others

ORIGIN

alignment\_scores:  
 Quality: 74.00 Length: 85  
 Ratio: 1.644 Gaps: 3  
 Percent Similarity: 52.941 Percent Identity: 25.882

alignment\_block:  
 US-09-142-524-1 x AU036117 ..

Align seg 1/1 to: AU036117 from: 1 to: 371

```

6 AlaPheAsnGlnPheGlyProAsnArgArgValPheIleLysArgValSe 22
:::||||| |||
83 AGTTCAACACATTC.....TTTCGTATACAC 108
:::|::| ||| ||||| |||
22 rAsnValIleIleHisGlyArgArgIleAspIlePheHisLysSer 39
:::|::| ||| ||||| |||
109 GAACACGAATATGATGAGCGTAGAAGATGCTACGCC..... 148
:::|::| ||| ||||| |||
39 heHisLeuGlnLysAsnThrIleGlyThrGlyArgArgIleSer..... 53
:::|::| ||| ||||| |||
149 .....ATGACTGCAGTCAAGGAGGATGGAGAAGATGACGAACCTTG 190
:::|::| ||| ||||| |||
54 .....LeuLysLeuThrSerGlyLysIleAl 62
:::|::| ||| ||||| |||
191 ATCTGCTTGAGCCGAGTGTGATCTCAGAAACGTGCCGAGAACTGAC 240
:::|::| ||| ||||| |||
62 aSerArgArgValAspGlyIleIleAlaIleAlaIleArgGlnAsnProAlaSer 79
:::|::| ||| ||||| |||
241 AGAAGATGAATCGACAGAGTGTCTACGATCTGCAAGAAATCTGCCCGCAGT 290
:::|::| ||| ||||| |||
79 rPLys 80
:::|::| ||| ||||| |||
291 ACAAG 295
:::|::| ||| ||||| |||

```

seq\_name: gb\_ests19:AA738025

seq\_documentation\_block:  
 LOCUS AA738025 520 bp mRNA EST 22-JAN-1998  
 DEFINITION nsl509.51 NCI CGAP GC3 Homo sapiens cDNA clone IMAGE:1256177 3'  
 similar to gb:X69150 40S RIBOSOMAL PROTEIN S18 (HUMAN);, mRNA

sequence.  
 ACCESSION AA738025  
 NID 92768782  
 VERSION AA738025.1 GI:2768782  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 520)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)

JOURNAL COMMENT

CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILM at:  
[www.dio.llnl.gov/bbrp/image/image.html](http://www.dio.llnl.gov/bbrp/image/image.html)

Insert Length: 579 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amerisham  
 High quality sequence stop: 402.

FEATURES  
 source location/Qualifiers  
 1..520  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="20q13.2-q13.3"  
 /clone\_lib="IMAGE:1256177"  
 /clone\_lib="NCI-CGAP\_GC3"  
 /tissue\_type="pooled germ cell tumors"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from 3 pooled  
 germ cell tumors, and was then primed with a Not I  
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
 RI adaptors (Pharmacia), digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT73  
 vector. Library is not normalized. Library was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 107 a 145 c 125 g 141 t 2 others

ORIGIN

alignment\_scores:  
 Quality: 75.00 Length: 79  
 Ratio: 1.829 Gaps: 2  
 Percent Similarity: 51.899 Percent Identity: 31.646

alignment\_block:  
 US-09-142-524-1 x AA738025/rev ..

Align seg 1/1 to reverse of: AA738025 from: 1 to: 520

```

12 ProAsnArgArgValPheIleLysArgValSerAsnValIleIleHisGI 28
|||::|::| ||| ||||| |||::|
485 CCGGAAGGTCACGACATATCTCGAGTAGACACACACCAACATCGATGG 436
|||::|::| ||| ||||| |||::|
28 yArgArgIleAspIlePheAlaSerLysAsnPhenHisLeuGlnLysAsn 45
|||::|::| ||| ||||| |||::|
435 TCGCGGAAATAGACCTTTCCTGACATGCTGCATTAAG..... 399
|||::|::| ||| ||||| |||::|
45 hrlIleGlyThrGlyArgArg..... 51
|||::|::| ||| ||||| |||::|
398 .....GGTGTGGCCGAGATATATGCTCATGTGTGTGGAAGACAGAC 354
|||::|::| ||| ||||| |||::|

```

```

52 11eserleuylsleuthrserglylslealaserargvalaspcl 68
||||:||||| .....:||||:
353 ATTGACTCACCAGAGAGCGGAGAACTCAGAGATGAGGTGACG 304
68 yllelelealalatyrglnaspralasertrplys 80
||||:||||| .....:|||||
303 TGTGATCACCATTATGACAGATCCAGCCAGTACAG 267

seq_name: gb_est29:AU036104

seq_documentation_block:
LOCUS AU036104 320 bp mRNA EST 26-MAR-1999
DEFINITION AU036104 Polyandrocampa misakiensis white spot budding stage
ACCESSION Polyandrocampa misakiensis cDNA, mRNA sequence.
NID AU036104
VERSION 94527065
KEYWORDS AU036104.1 GI:4527065
SOURCE EST.
ORGANISM Polyandrocampa misakiensis.
Eukaryota; Metazoa; Chordata; Trochordata; Ascidiacea;
Stolidobranchia; Scyliidae; Polyandrocampa.
REFERENCE
1 (bases 1 to 320)
Kawamura, K., Hayata, D., Fujiwara, S. and Yubisui, T.
Serine protease inhibitors expressed in the process of budding of
tunicates as revealed by EST analysis
J. Biochem. (1998) In press
On Jun 5, 1998 this sequence version replaced gi:3188155.

JOURNAL
COMMENT

Contact: Kawamura K
Faculty of Science
Kochi University
2-5-1, Akebono-cho, Kochi 780, Japan
Tel: +81-888-44-8313
Fax: +81-888-44-8313
Email: kazuk@cc.kochi-u.ac.jp.
Location/Qualifiers
1. 320
/organism="Polyandrocampa misakiensis"
/strain="white spot"
/db_xref="taxon:7723"
/clone_id="Polyandrocampa misakiensis white spot budding
stage"
/dev_stage="budding stage"

BASE COUNT 96 a 70 c 85 g 68 t 1 others
ORIGIN

alignment_scores:
Quality: 73.00 Length: 78
Ratio: 1.659 Gaps: 2
Percent Similarity: 56.410 Percent Identity: 25.641

alignment_block:
US-09-142-524-1 x AU036104 ..

Align seg 1/1 to: AU036104 from: 1 to: 320

1 4 Prosmnargargvalpellelyargvalaserasnvallelelscg 28
||||:||||| .....:||||:
68 CCAAGAGAGTTTCAACACATCTTCGTATCAGAACGCAATTCGATG 117
32 yargarglleaspllephelaserlysasnphelhsleuglnlyasnt 45
||||:||||| .....:|||||
118 ACAGTGAAGATCTCTACGCC.....ATGACTGACG 149
45 hrtleeglythrglyargargileser..... 53
150 TCAAGGAGAGTTGAGAGATTCAGCACTGATTCGAAGAAGCGCGAT 199
54 .....leuylsleuthrserglylslealaserargvalaspcl 68
200 GTTATCTCAGGAAGCGTCCGAGAGACGACGAGATGAATCGACAG 249

```

```

68 yllelelealalatyrglnaspralasertrplys 79
||||:||||| .....:|||||
250 AGTGTGACAGATCAGATCAGAACTCAGAGATGAGGTGACG 283

seq_name: gb_est12:AA300566

seq_documentation_block:
LOCUS AA300566 265 bp mRNA EST 18-APR-1997
DEFINITION EST13420 Testis tumor Homo sapiens cDNA 5' end similar to similar
to ribosomal protein S18, mRNA sequence.
ACCESSION AA300566
NID q1952919
VERSION AA300566.1 GI:1952919
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 265)
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,
Balt, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-ai, C.,
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,
Fitzgerald, L.M., Fitch, W.M., Fitchman, V.L., Geoghegan, N.S.,
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
Kelley, T.M., Kelley, J.C., Liu, L.-T., Marmaros, S.M., Merrick, J.M.,
Moreno-Palancas, R.F., McDonald, J.L., Nguyen, D.T., Peligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudak, D.M., Shiley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Fertle, A., Fischer, C., Hastings, G.A.,
He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,
Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H.,
Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M.,
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
Fraser, C.M., and Venter, J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280

JOURNAL
MEDLINE
COMMENT
On Sep 12, 1996 this sequence version replaced gi:1395394.

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1. 265
/organism="Homo sapiens"
/db_xref="ATCC (inhost):192428"
/db_xref="taxon:9606"
/clone_id="Testis tumor"
/sex="male"
/dev_stage="adult"
/note="Organ: testis; Vector: pbluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

BASE COUNT 70 a 61 c 62 t 4 others
ORIGIN

alignment_scores:
Quality: 72.00 Length: 79
Ratio: 1.756 Gaps: 2
Percent Similarity: 51.899 Percent Identity: 30.380

```

alignment\_block:  
US-09-142-524-1 x AA300566 ..

Align seg 1/1 to: AA300566 from: 1 to: 265

```

12  ProAsnArgArgValPheIleLysArgValSerAsnValIleIleHisG1 28
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30  CCGAAAGATCCAGCATTTTTCGAGTACGACACACACATCGATGG 79
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28  YARGArgIleAspIlePheAlaSerLysAsnPhenHisLeuGlnLysAsnT 45
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
80  GCGGCGGAAATAGCCTTCCATCCTACCTCCATTAAG..... 116
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45  hTleGlyThrGlyArgArg..... 51
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117  ....GCTGGGGCGAGATATGCTCATGTGTTGAGCAACACAC 161
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52  IleSerLeuLysLeuThrSerGlyLysIleAlaSerArgArgValAspG1 68
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
162  ATTGACTCTACCAAGAGGGGGGAGAGACCTACAGAGNTGAGGTGAACG 211
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68  YIleIleAlaIleArgGlnAsnProAlaSerTplys 80
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
212  TGTGATCACCATTATTCAGATCCAGCAGTACAAG 248
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq\_name: gb\_est9:D76822

seq\_documentation\_block: 308 bp mRNA EST 07-OCT-1997  
LOCUS D76822  
DEFINITION MUS78A01 mouse embryonal carcinoma cell line F9 Mus musculus cDNA  
clone 78A01, mRNA sequence.

ACCESSION D76822  
NID D76822  
VERSION D76822.1 GI:1596557  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 308)  
AUTHORS Nishiguchi, S., Sakuma, R., Nomura, M., Zou, Z., Jeonatsilavong, J.,  
Joh, T., Yasunaga, T. and Shimada, K.  
TITLE A catalogue of genes in mouse embryonal carcinoma F9 cells  
identified with expressed sequence tags

JOURNAL J. Biochem. 119, 749-767 (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1405026.

CONTACT: Kazunori Shimada  
Department of Medical Genetics, Division of Molecular Biomedicine  
Research Institute for Microbial Diseases, Osaka University  
3-1, Yamadaoka, Suita, Osaka, 565, Japan  
Tel: 06-879-8325  
Fax: 06-879-8326.

FEATURES  
SOURCE location/Qualifiers

1..308  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_id="78A01"  
/clone\_lib="mouse embryonal carcinoma cell line F9"

RASP COUNT 82 a 74 c 91 g 57 t 4 others

alignment\_scores:

Quality: 72.00 Length: 79  
Ratio: 1.756 Gaps: 2  
Percent Similarity: 51.899 Percent Identity: 30.380

alignment\_block:

US-09-142-524-1 x D76822 ..  
Align seg 1/1 to: D76822 from: 1 to: 308

seq\_name: gb\_est17:AA603354

seq\_documentation\_block: 583 bp mRNA EST 08-OCT-1997  
LOCUS AA603354  
DEFINITION np06a11.s1 NCI-CGAP\_P3 Homo sapiens cDNA clone IMAGE:1115516  
similar to gb:X69150 40S RIBOSOMAL PROTEIN S18 (HUMAN); mRNA  
sequence.

ACCESSION AA603354  
NID 92437215  
VERSION AA603354.1 GI:2437215  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 583)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1394196.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linhan, M.D., Rodrigo Chuagui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.llnl.gov/dbp/limage/limage.html

Insert Length: 635 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from AmerSham  
High quality sequence stop: 437.

FEATURES  
SOURCE location/Qualifiers

1..583  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="8"  
/clone="IMAGE:1115516"  
/clone\_lib="NCI-CGAP\_P3"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: PAMP10. Site 1: NotI. Site 2: EcoRI. 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected cells  
histologically determined to be fully malignant prostate  
cancer cells. Double-stranded cDNA was ligated to EcoRI

adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the U93-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman.

BASE COUNT 118 a 161 c 130 g 174 t  
ORIGIN

## alignment\_scores:

Quality: 74.00 Length: 79  
Ratio: 1.805 Gaps: 2  
Percent Similarity: 51.899 Percent Identity: 31.646

## alignment\_block:

US-09-142-524-1 x AA603354/rev ..

Align seg 1/1 to reverse of: AA603354 from: 1 to: 583

```

12 ProsaAARGValPheIleLysArgValSerAsnValIleHISGI 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
556 CCTGAAGAAGTCCAGCATATTGCGAGTACTCAACACCAACATCGATGG 507
28 YARGArgIleAspIlePheAlaSerLysAsnPheHisLeuGlnLysAsnT 45
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
506 GCGGCGGAAATAGCCTTGCCATCTCATCTGCATTAAG..... 470
45 hrIleGlyThrGlyArgArg..... 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 .....GCTGTGGCCGAGATATGCTATGCTGTGGAGAAACAGAC 425
52 IleSerLeuLysLeuThrSerGlyLysIleAlaSerArgValAspGI 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 ATGACCTCACCAGAGGCGGAGGAGACTCGCTGAGAGTGGTGAACG 375
68 YIleIleAlaIleAlaTyrlGlnAsnProAlaSerTrpLys 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 TGTGATCACCATTATGACGATTCACGCCAGTACAG 338

```

seq\_name: gb\_est13:AA372511

## seq\_documentation\_block:

LOCUS AA372511 343 bp mRNA EST 21-APR-1997  
DEFINITION EST84502 Colon adenocarcinoma IV Homo sapiens cDNA 5' end similar to similar to ribosomal protein S18, mRNA sequence.  
ACCESSION AA372511  
NID 92024904  
VERSION AA372511.1 GI:2024904  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 343)  
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulton,R.A., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wei,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.T., Maturo,S.M., Merrick,J.M., Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Ulterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dime,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M., and Venter,J.C.

TITLE Initial assessment of human gene diversity and expression patterns

JOURNAL  
MEDLINE  
COMMENT based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
96026280  
On May 8, 1995 this sequence version replaced gi:800909.

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlavage@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

## FEATURES

## source

Location/Qualifiers

1..343

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):176944"

/db\_xref="taxon:9606"

/clone\_lib="Colon adenocarcinoma IV"

/dev\_stage="adult"

/note="Organ: colon; Vector: pBluescript SK-; Site:1: EcoRI; Site:2: XhoI"

BASE COUNT 99 a 75 c 90 g 74 t 5 others

## alignment\_scores:

Quality: 72.00 Length: 79  
Ratio: 1.714 Gaps: 2  
Percent Similarity: 53.165 Percent Identity: 30.380

## alignment\_block:

US-09-142-524-1 x AA372511 ..

Align seg 1/1 to: AA372511 from: 1 to: 343

```

12 ProsaAARGValPheIleLysArgValSerAsnValIleHISGI 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 CCTGAAGAAGTCCAGCATATTGCGAGTACTCAACACCAACATCGATGG 81
28 YARGArgIleAspIlePheAlaSerLysAsnPheHisLeuGlnLysAsnT 45
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
82 GCGGCGGAAATAGCCTTGCCATCTCATCTGCATTAAG..... 118
45 hrIleGlyThrGlyArgArg..... 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 .....GCTGTGGCCGAGATATGCTATGCTGTGGAGAAACAGAC 163
52 IleSerLeuLysLeuThrSerGlyLysIleAlaSerArgValAspGI 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 ATGACCTCACCAGAGGCGGAGGAGACTCGCTGAGAGTGGTGAACG 213
68 YIleIleAlaIleAlaTyrlGlnAsnProAlaSerTrpLys 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 TGTGATCACCATTATGACGATTCACGCCAGTACAG 250

```

seq\_name: gb\_est14:AA381359

## seq\_documentation\_block:

LOCUS AA381359 263 bp mRNA EST 21-APR-1997  
DEFINITION EST94431 Activated T-cells I Homo sapiens cDNA 5' end similar to similar to ribosomal protein S18, mRNA sequence.  
ACCESSION AA381359  
NID 92033679  
VERSION AA381359.1 GI:2033679  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS

1 (bases 1 to 263)

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, W.A., Cline, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dime, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)

On Sep 12, 1996 this sequence version replaced gi:1405101.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tldb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
source

1. 263  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):185717"  
/db\_xref="taxon:9606"  
/clone\_lib="Activated T-cells I"  
/cell\_type="T-lymphocyte"  
/dev\_stage="adult"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 71 a 62 c 70 g 57 t 3 others  
ORIGIN

alignment\_scores:  
Quality: 71.00 Length: 79  
Ratio: 1.732 Gaps: 2  
Percent Similarity: 51.899 Percent Identity: 30.380

alignment\_block:  
US-09-142-524-1 x AA381359 ..

Align seg 1/1 to: AA381359 from: 1 to: 263

```

12  PROSAAAGAGValpHeilelyArGValserAsnVallelleHiegl 28
   |||||..... ||| ||||| ||||| ||||| |||||
43  CCGTGAAGAGTTCACGATATTTCAGAGTACACACCAACATGATGATG 92
   |||||..... ||||| ||||| ||||| |||||
28  YAGAGGIIleAspIIlePheAlaSerLysAsnPhelHisleuGlnLysAsn 45
   |||||..... ||||| ||||| ||||| |||||
93  CGCGCGAAATAGACCTTCCTCCATCCTCCATTAAG..... 129
   |||||..... ||||| ||||| ||||| |||||
45  hrIleGlyThrGlyArgArg..... 51
   |||||..... ||||| ||||| ||||| |||||
130  ....GCTGTGGCGCGAATATGCTCATGTGTGTGAGGAAGACGAC 174
   |||||..... ||||| ||||| ||||| |||||
52  IlleSerLeuLysLeuThrSerGlyLysIleAlaSerArgValAspG1 68
   |||||..... ||||| ||||| ||||| |||||

```

175 APTGACCTCACCAGAGGCGGAGAACTCAGAGATGAGGTGGAACG 224

68 YIleAlaAlaIleAlaTrpGlnAsnProAlaSerTrpLys 80

225 TGTGTACCACTATATGCAATCCACGCGATGACG 261

seq\_name: gb\_est13:AA379609

seq\_documentation\_block:

LOCUS AA379609

DEFINITION EST92467 SK1n tumor I Homo sapiens cDNA 5' end similar to similar

to ribosomal protein S18, mRNA sequence.

ACCESSION AA379609

VERSION 20031927

KEYWORDS AA379609.1 GI:2031927

SOURCE EST

ORGANISM human

REFERENCE

AUTHORS

1 (bases 1 to 268)  
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, W.A., Cline, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dime, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)

On Sep 12, 1996 this sequence version replaced gi:1394110.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tldb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
source

1. 268  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):183986"  
/db\_xref="taxon:9606"  
/clone\_lib="Skin tumor I"  
/dev\_stage="adult"  
/note="Organ: skin; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 72 a 65 c 71 g 57 t 3 others  
ORIGIN

alignment\_scores:  
Quality: 71.00 Length: 79  
Ratio: 1.732 Gaps: 2  
Percent Similarity: 51.899 Percent Identity: 30.380



TITLE Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)  
 MEDLINE 96026280  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1395403.

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/cdb/hgi/hgi.html>)

Seq primer: M13 Reverse

Location/Qualifiers

1..287

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):182414"

/db\_xref="taxon:9606"

/map="809805:18:18p11.21-18p11.23:17q21"

/clone.lib="Synovial sarcoma"

/sex="female"

/tissue\_type="synovial membrane"

/dev\_stage="adult, 20 yrs"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

BASE COUNT 78 a 68 c 77 g 60 t 4 others

ORIGIN

alignment\_scores:

Quality: 71.00 Length: 79  
 Ratio: 1.732 Gaps: 2

Percent Similarity: 51.899 Percent Identity: 30.380

alignment\_block:

US-09-142-524-1 x AA378031 ..

Align seg 1/1 to: AA378031 from: 1 to: 287

```

12 ProasnaargavAlpneileysargValserAsnValleleHscl 28
|||||:|||||
38 CCTGAAAGTCCAGCATATTTCGAGTACTCAACACACATCGATGg 87
|||||
28 YArgArgTlleaspIlehealaserLysasnphenisLeugInLysasnT 45
|||||
88 GCGGCGGAAATAGCCTTGGCCATCACTGCCATTAAg..... 124
|||||
45 hrlleGlyThrGlyArgArg..... 51
|||||
125 .....GGTGTGGCCGAGCATATGCTCATGTGTGTGAGGAAGACAGAC 169
|||||
52 lleSerleuLysLeuThrSerGlyLysIlealaserargValaspGl 68
|||||
170 ATTGACCTCACCAGAGGGCGGAGAGACTCACTGAGGATGAGGTGGAACG 219
|||||
68 yllellelealalaIatyrGlnasnProalaserTrpLys 80
|||||
220 TGTGATCACCATTATGCAGATCCAGCCAGTACAG 256

```